

## Data trustworthiness and user reputation as indicators of VGI quality

Paolo Fogliaroni, Fausto D'Antonio & Eliseo Clementini

To cite this article: Paolo Fogliaroni, Fausto D'Antonio & Eliseo Clementini (2018) Data trustworthiness and user reputation as indicators of VGI quality, Geo-spatial Information Science, 21:3, 213-233, DOI: [10.1080/10095020.2018.1496556](https://doi.org/10.1080/10095020.2018.1496556)

To link to this article: <https://doi.org/10.1080/10095020.2018.1496556>



© 2018 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 21 Sep 2018.



Submit your article to this journal [↗](#)



View Crossmark data [↗](#)

## Data trustworthiness and user reputation as indicators of VGI quality

Paolo Fogliaroni <sup>a</sup>, Fausto D'Antonio<sup>b</sup> and Eliseo Clementini <sup>b</sup>

<sup>a</sup>Department for Geodesy and Geoinformation, Vienna University of Technology, Vienna, Austria; <sup>b</sup>Department of Industrial and Information Engineering and Economics, University of L'Aquila, L'Aquila, Italy

### ABSTRACT

Volunteered geographic information (VGI) has entered a phase where there are both a substantial amount of crowdsourced information available and a big interest in using it by organizations. But the issue of deciding the quality of VGI without resorting to a comparison with authoritative data remains an open challenge. This article first formulates the problem of quality assessment of VGI data. Then presents a model to measure trustworthiness of information and reputation of contributors by analyzing geometric, qualitative, and semantic aspects of edits over time. An implementation of the model is running on a small data-set for a preliminary empirical validation. The results indicate that the computed trustworthiness provides a valid approximation of VGI quality.

### ARTICLE HISTORY

Received 21 March 2018  
Accepted 4 June 2018

### KEYWORDS

Volunteered geographic information; data quality; trustworthiness; reputation

## 1. Introduction

Volunteered geographic information (VGI) (Goodchild 2007) is a form of user-generated content (UGC) that is primarily concerned with the survey, collection, and dissemination of spatial data. Recently, VGI projects gained increasing attention and, as a consequence, their user basin increased notably. Several studies showed that while VGI coverage and quality is approaching that of authoritative data-sets (in areas with large number of volunteers), VGI quality is not homogeneously distributed in space (Haklay 2010; Mooney, Corcoran, and Winstanley 2010). Accordingly, the conception and implementation of new methods to assess VGI quality is highly needed (Elwood, Goodchild, and Sui 2013).

In general, data quality can be regarded as the level of fitness between single pieces of data and the parts of reality that these represent – i.e. the more precisely data corresponds to reality, the higher its quality.

Thus, the only way to obtain the “true” quality value of data would be to compare it to reality, but this is impracticable because data and reality belong to two different spaces (conceptual and physical, respectively). In practice, data quality is typically evaluated with respect to four main dimensions: accuracy, completeness, consistency, and timeliness (Wand and Wang 1996).

There are two main approaches to assess VGI data quality. The first one assumes that authoritative data are quality data. Thus, it consists in comparing volunteered against authoritative data-sets and addresses three of the four dimensions mentioned above: accuracy, completeness, and consistency. One

main drawback of this approach is that it requires the access to authoritative data, which might not be possible because of limited data availability, licensing restrictions, or high procurement costs (Antoniou and Skopeliti 2015). Moreover, while authoritative data are regarded as quality data, in fact there is no absolute guarantee about its correctness and consistency, especially considering the discrepancies between data and reality that arise over time because of the slow update rate of authoritative data-sets. The second approach aims at assessing the quality of VGI by analyzing the evolution of the data itself, i.e. its history or provenance. This approach overcomes the limitations of the previous one as it does not require external data sources and can also take into account the timeliness dimension. However, this approach provides an approximation of the quality, rather than an accurate measurement. In other words, it provides a proxy measure for data quality that is referred to as trustworthiness (Dai et al. 2008).

This article presents a system to compute a trustworthiness score for each version of a geographic feature by analyzing its provenance and evolution. Typically, in the literature, provenance is defined as the historical evolution of data over time, but in the scope of this work we distinguish between data provenance and data evolution. Fixed a point in time, by provenance we intend the sequence of edits that a feature undergoes over time until the given time instant; by evolution we intend the edits that the feature will undergo in the future. The edit sequence of a feature over time is assumed to be implemented through versioning. Both provenance and evolution

include authoring information. Accordingly, our system computes two scores: (1) *reputation* is a score that is associated to a volunteer and that denotes her reliability; (2) *trustworthiness* is a score that is associated to each version of a geographic feature and that depends on both the sequence of edits that the feature undergoes over time and the reputation of its author.

The main underlying idea is that, according to the many eyes principle (Raymond 1999), the quality of a feature will improve over time and editing: the more volunteers contribute information about a feature, the higher the probability that errors and inconsistencies are spotted and corrected, and, thus, the higher the probability that the feature is correctly mapped in the VGI system. Such a concept is captured by the notion of confirmation. Approaches to encourage crowdsourcing take into account people's intrinsic motivation, such as in Martella, Kray, and Clementini (2015).

We distinguish three main effects making up a confirmation: direct, indirect, and temporal. The direct effect corresponds to the case where a new version of a feature is contributed that leaves untouched a subset of the attributes of the current feature version. We assume that this is an indication that the attribute values were already correct, thus the overall trustworthiness of the feature must increase. The indirect effect covers the cases where a feature is edited that is close enough to the feature being confirmed. Since the volunteer only modified the nearby feature, this can be considered as an indication of the quality of the feature being confirmed. Finally, the temporal effect addresses the persistency of a feature version over time, i.e. the longer a version stays untouched, the higher the probability that it is correctly mapped.

Direct and indirect effects are finer divided into three components, to more precisely account for the different aspects characterizing the editing preferences of contributors: *thematic*, *geometric*, and *qualitative*. The first two components have been included in order to model more precisely the skills and the preferences of the volunteers. For example, a volunteer with little or no surveying skills may be more inclined to contribute information only about the thematic component of a geographic feature and leave untouched the geometric part. This means that confirmations generated by these kinds of people must be weighted differently for the geometric and the thematic aspects.

The qualitative aspect has been introduced to address properties of human spatial cognition. Findings in cognitive science (Mark 1993) have shown that humans conceptualize space in a relative and qualitative manner, i.e. mental representations of space are typically geometrically distorted, but some

qualitative relations among objects (e.g. order, direction, and topology) are interiorized correctly. This means that while a volunteer may be unable to notice a small geometric imprecision in a piece of data, she/he may notice much more easily a wrong qualitative spatial relation (e.g. two disconnected buildings that have been mapped as touching each other).

The rest of the article is structured as follows. In Section 2, we discuss more in detail background work on the subject. Section 3 is the core of our model: we discuss the types of changes that affect VGI data, we define trustworthiness and reputation as the main aspects that influence VGI data quality, and we distinguish among direct, indirect, and temporal effects. In Section 4, we describe a system architecture that implements our model: the resulting framework, called TandR (for Trustworthiness and Reputation), makes use of domain ontologies for data representation. In Section 5, we carried out an experiment with a selected portion of OpenStreetMap (OSM) data (<http://www.openstreetmap.org>), whose various feature versions were compared to ground truth data to explore how trustworthiness and reputation indexes vary over time following data editing. Section 6 draws short conclusions and highlights future improvements and extensions.

## 2. State of the art

While VGI is a very successful means to collect geographic information at low cost (Goodchild 2007), it suffers a major drawback: VGI comes with no assurance of quality (Goodchild and Li 2012). The issue is inherently related to the very nature of VGI, that is provided by volunteers in a relatively unconstrained manner. Indeed, most of the VGI systems currently implemented do not impose any enforcement on their contributors, except for some guidelines to streamline the input data. So, according to the mainstream philosophy of VGI, there is no control in place that guarantees the goodness of the contributed information.

Recent surveys (Antoniou and Skopeliti 2015; Senaratne et al. 2017; Degrossi et al. 2018; Fonte et al. 2017) perform a review of quality measures and indicators for VGI. Specifically, Senaratne et al. (2017) discriminate among three different forms of VGI systems: map-based (e.g. OSM), image-based (e.g. Flickr), and text-based (e.g. geo-blogs). They conclude that the first form is by far the most widespread. As quality measures for map-based VGI, they identify completeness, consistency, positional accuracy, temporal accuracy, and thematic accuracy. These quality measures are defined in ISO 19157. As for quality indicators, they find trustworthiness, credibility, and reputation, among others.

Accordingly, there exists two main methods to assess VGI quality. The first compares VGI data-sets against professionally surveyed ground truth data-sets. In this case, it is possible to assess the aforementioned quality measures. The second aims at deriving some quality indicators only by analyzing the VGI data-sets themselves, without a comparison with external sources.

Several studies have carried out in the last years to assess quality of map-based VGI (mainly OSM) by comparison against professionally surveyed data-sets. Most studies focused on quality assessment of road networks in England (Haklay 2010), Germany (Zielstra and Zipf 2010; Neis, Zielstra, and Zipf 2012), and France (Girres and Touya 2010), among other countries. Other work focused on general features rather than street networks only (Helbich et al. 2012; Mooney, Corcoran, and Winstanley 2010; Fan et al. 2014). All such contributions came to similar conclusions: (1) coverage and accuracy of map-based VGI data-sets are approaching those of professionally surveyed data-sets and (2) VGI quality is not homogeneously distributed, with peaks in the most populated areas.

Goodchild and Li (2012) suggest three approaches to assure VGI quality without resorting to external data sources. The first is termed “crowdsourcing” and assumes that quality increases with the number of contributors. Basically, this is an adaptation of the so-called Linus’s Law in honor of Linus Torwalds. Speaking about open-source software, Torwalds stated that “given enough eyeballs, all bugs are shallow” (Raymond 1999). When adapted to crowdsourcing, Linus’s Law is sometimes referred to as the many eyes principle: “If something is visible to many people then, collectively, they are more likely to find errors in it. Publishing open data can therefore be a way to improve its accuracy and data quality, especially where a good interface for reporting errors is provided” (<http://opendatahandbook.org/glossary/en/terms/many-eyes-principle/>). Therefore, Goodchild and Li (2012) conclude that the crowdsourcing approach does not properly apply to less known facts, such as geographic features located in a sparsely populated area. Indeed, in this case, the many eyes needed to assure the quality would be missing. The second approach mentioned by Goodchild and Li (2012) was named “social”. It relies on the construction of a hierarchy of moderators, i.e. individuals that are reputed trustworthy in the community because of the quality of their contributions. The third approach is termed “geographic” and is the one that is best suited for full or semiautomatization: it decides on the quality of a feature by comparing it against

geographical laws – e.g. a shoreline should have a fractal shape.

The many eyes principle comes in support of collective intelligence. According to this theory, a group of individuals performs better than its best members. Spielman (2014) analyzes the issue of VGI quality from this perspective. He argues that there are two approaches to quality assessment: validation-by-accuracy and validation-by-credibility. The former corresponds to measuring VGI quality by comparison against ground truth data-sets. The latter aims at deriving VGI quality indicators based on the credibility of the volunteer contributing a feature that, in turn, depends on reputation, trustworthiness, and motivation of the volunteer (Bishr and Mantelas 2008). Spielman (2014) concludes that VGI systems should be designed to foster collective intelligence.

### **2.1. VGI quality indicators: trustworthiness and reputation**

Bishr and Janowicz (2010), among others, suggest to use data trustworthiness (aka reliability) as a proxy measure for VGI quality. Arguably, the reliability of an object, a person, or an action corresponds to its grade of predictability. Reliability reflects this idea and corresponds to “a bet about the future contingent actions of others” (Sztompka 1999). Mezzetti (2004) provides a compatible definition of trustworthiness by asserting that an entity is trustworthy, within a given context, if it actually justifies reliance on the dependability of its behavior within that context. Additionally, Bishr and Janowicz (2010) stress the notion of people-object transitivity: the degree of reliability associated to a person propagates to the entities that are somehow connected to her/him. Moreover, they argue that the reputation of a person indicates how much this person is considered reliable within a community.

Reliability indicators relating to spatial features contributed in a VGI system (trustworthiness) and to their contributors (reputation) seem to be valid approximations of VGI data quality. Lodigiani and Melchiori (2016) propose an adaptation of the PageRank algorithm for web pages to derive contributors reputation. Barron, Neis, and Zipf (2014) introduce a framework that incorporates 25 quality indicators and corresponding methods for OSM data quality. Finally, Keßler, Trame, and Kauppinen (2011) propose to derive reliability scores from historical information of VGI items. They argue that, from the edit history of features, one can obtain all necessary information to define trustworthiness scores for the spatial features in a VGI system and reputation scores for their contributors.

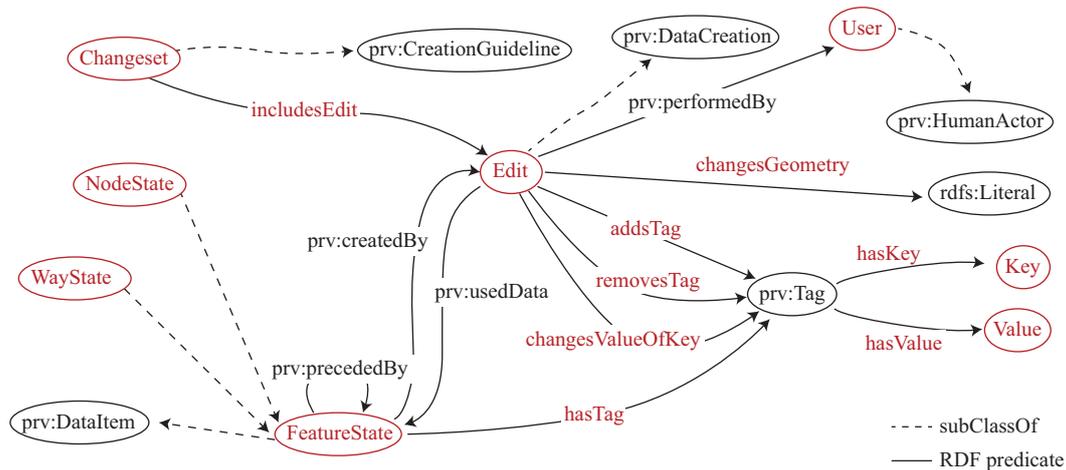


Figure 1. OSM provenance ontology diagram.

## 2.2. Data provenance

In order to use historical information to derive quality indicators, it is convenient to use an appropriate model. Keßler, Trame, and Kauppinen (2011) tackle the challenge by adapting the concept of data provenance (Hartig 2009) to VGI. More specifically, Keßler, Trame, and Kauppinen (2011) introduce the ontology in Figure 1 to represent OSM data provenance. They introduce the concept of “Editing Pattern”: a sequence of editing actions from which it is possible to deduce useful information to derive a score for user reputation and data trustworthiness. The identified patterns are “Confirmations”, “Corrections”, and “Rollbacks”, including the special cases of “SelfCorrections” and “SelfRollbacks”. Each editing pattern has a different effect on reputation and trustworthiness. Keßler and Groot (2013) suggest that the trustworthiness of a feature should grow proportionally with the number of versions, contributors, and confirmations, and inversely with the number of corrections and rollbacks.

## 3. Modeling trustworthiness and reputation

This section details a novel model to derive data trustworthiness and user reputation from feature edit sequences. A first version of the model appeared in D’Antonio, Fogliaroni, and Kauppinen (2014). We drew inspiration from the work of Keßler, Trame, and Kauppinen (2011) and Keßler and Groot (2013) – see Section 2.2 – but our model differs from other approaches with regard to the following aspects. Our model:

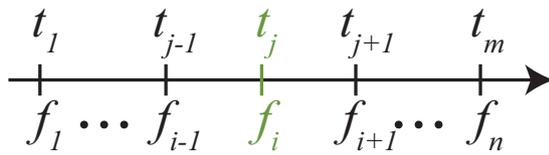
- (1) can fit any map-based VGI system (rather than OSM only) provided that the system implements feature versioning and that differences between versions are ascribable to series of atomic operations; namely, creation, modification, and deletion;

- (2) associates a reputation score to each VGI author;
- (3) derives a trustworthiness score for each feature version (rather than for each feature);
- (4) accounts for the impact of changes between two versions according to their relevance;
- (5) derives trustworthiness and reputation scores also considering evolution information (i.e. future edits, rather than only provenance information).

## 3.1. Model overview

Our first goal was to design a model to derive trustworthiness and reputation scores from both data provenance and evolution. That is, rather than evaluating a feature contributed at a specific point in time only by analyzing current and historical information in the system, we also exploit information that has to come. For this to be possible, two conditions must be satisfied: (1) the VGI system at hand implements versioning; (2) every time an event alters the state of the information system, the trustworthiness score of features affected directly or indirectly (see Section 3.5) by the change are also updated.

Versioning is a method that prevents data from being lost and defines a temporal order among feature states (as shown in Figure 2). The state of a feature  $f$  is called feature version and is, simply, a set of attributes describing the feature at a given point in time. We denote by  $f_i$  the  $i$ -th version of a feature  $f$ . A new version is established every time a user of the VGI system creates or deletes a feature, or modifies a non-empty subset of the attributes of an existing feature. This user is the author of the version and to each version is associated exactly one author. The time interval between a feature version and the next is called the version lifetime. The creation of a feature starts a feature version lineage. Each new version extends the lineage, which terminates with the deletion of the feature.



**Figure 2.** Versioning establishes a temporal ordering on feature states.

### 3.2. Model dynamics

Every time a new feature version  $f_i$  is added to the VGI system the feature version lineage of  $f$  gets extended and we use the new bit of information to update trustworthiness and reputation for a subset of the system's features and authors. For example, assume that feature  $f$ , currently at version  $f_i$ , is modified, generating version  $f_{i+1}$ . The model is designed to derive a trustworthiness value for the new feature version by comparing it with previous versions. Concurrently, the reputation of the author of the new feature will also be updated. The new version also triggers an adjustment of the trustworthiness of previous versions of  $f$ , as now we have a new version that they can be compared against. Consequently, the reputation of the authors of these versions will also be adjusted. Finally, the newly introduced version indirectly affects the trustworthiness of currently alive versions of “nearby” features (see Section 3.5 for details) and the reputation of their authors.

Each time a trustworthiness or reputation score is updated, it is assigned a *validity* timestamp, which denotes that the score value accounts for all information available in the VGI system from that point in time; it will be valid until the next update. Also, in order to support the analysis and the monitoring of the system dynamics, scores are never deleted: simply, newer scores supersede older ones.

### 3.3. Types of edits

We call edit any operation on the VGI system that generates a new feature version. Some previous approaches (Keßler, Trame, and Kauppinen 2011; Keßler and Groot 2013) focus on specific VGI systems (i.e. OSM) and aims at modeling emergent editing patterns of the system at hand. Such an approach allows for treating more accurately specific patterns: one exemplary editing pattern goes under the name of “edit wars” (Keßler and Groot 2013) or “tag wars” (Mooney and Corcoran 2012); two or more users keep changing the attributes of a feature back and forth between the same states. Nevertheless, the approach delivers a rigid model, in the sense that it hardly applies to VGI systems exposing different editing patterns than the one for which it has been

designed. Conversely, we favored generality and flexibility over specialty and designed a model to respond to atomic editing operations.

#### 3.3.1. Creation

A new feature is added to the VGI system, along with its first version. Creation starts a version lineage that is used at later points in time to derive trustworthiness scores. At creation time, though, there is no historical information available to derive the trustworthiness of the new feature version, so this is set equal to the reputation of its author.

#### 3.3.2. Modification

A new feature version is generated by adding, altering, or deleting some attributes of the latest version of the feature. At modification time, the trustworthiness of the new version is derived and the trustworthiness of previous versions is updated according to the similarity to all versions in the lineage, which was just extended by the new version.

#### 3.3.3. Deletion

A feature is removed from the VGI system. In fact, this creates a special version of the feature denoting the end of the feature lineage. The trustworthiness of this version is set equal to the reputation of its author and the trustworthiness of previous versions are not updated. Indeed, a deletion may denote one of two things: either the feature is not existing any longer in the reality or the information provided is totally wrong. In the first case, the trustworthiness of previous versions should not be affected as these were established in the past, when the feature was still existing. In the latter case, according to the many eyes principle, the wrong information will eventually be corrected by restoring the deleted feature. Restoration is treated as a modification, meaning that the trustworthiness associated to the “deleted” version will decrease drastically and, proportionally, the reputation of its author.

### 3.4. Impact and aspects of edits

In general, the trustworthiness of a feature at a given point in time is a function of the similarity among the versions in the feature lineage. We adopt the typical geographical information system (GIS) perspective that considers a geographic feature as consisting of thematic and geometric attributes, but we further add a qualitative spatial level that takes into account more drastic changes that might be caused by small geometric changes. In fact, a small geometric change may change drastically the qualitative spatial relations holding among a feature and its neighbors and vice versa.

Accordingly, for the computation of trustworthiness and reputation, we consider the aspects reported in the following subsections.

### 3.4.1. Thematic aspect

The thematic aspect describes the nongeometric characteristics of the feature. For example, the fact that the feature represents either a road, a traffic light, or a building with an orange facade. Thematic information is assumed to be represented by key–value pairs.

A simple approach to account for thematic edits is to look at the number of differences between the pairs associated to the new version and the pairs associated to other versions in the lineage. A finer-grained alternative would be to look at the semantic similarity between the values associated to the different versions. For example, this can be done by considering the lexical distance (i.e. the shortest path) between two terms on the Wordnet graph (<http://graphwords.com/>) (Miller 1995). Alternatively, the shortest distance in an ontology including both previous and altered key–value pairs can be used.

For example, consider the case where the key “amenity” of feature version  $f_i$ , initially set to “Establishment”, undergoes two thematic edits. The first edit changes its value to “University”, and the second to “Educational Institution”. If we only look at the number of thematic differences, both edits would result in a score of one. Conversely, using the lexical distance on the Wordnet graph (as shown in Figure 3), it would be possible to better characterize the edits: the first edit would be characterized by a score of three, the latter by a score of six with respect to  $f_i$ .

### 3.4.2. Geometric aspect

The geometric aspect describes shape and position of a feature. The relevance of a geometric edit is



Figure 3. Wordnet graph example.

evaluated with respect to a series of geometrical properties, such as perimeter, area, number, and position of vertices. Versions of a feature (as shown in Figure 4) are compared to see the differences in these geometric components. For example, version  $f_j$  has more vertices than version  $f_i$ , but lesser area and perimeter.

### 3.4.3. Qualitative spatial aspect

The qualitative spatial aspect refers to changes in the qualitative spatial relations holding between a feature and its neighbors. According to Clementini (2013), qualitative spatial relations belong to one of three classes: topological (Egenhofer 1989; Randell and Cohn 1989; Tarquini and Clementini 2008), projective (Clementini et al. 2010; Tarquini et al. 2007; Fogliaroni and Clementini 2015; Billen and Clementini 2006, 2005b, 2005a), and metric (Hernández, Clementini, and Di Felice 1995; Moratz and Ragni 2008).

The simultaneous consideration of both the geometric and qualitative spatial aspects allows for weighting more appropriately spatial edits. A qualitative spatial change happens when a modification to the geometric properties of a feature alters at least one qualitative spatial relation between the feature and its neighbors. The impact on trustworthiness is a function of the distance on the conceptual neighborhood graph between the relations holding before and after the qualitative spatial change.

As an example, let us consider topological relations only. Consider the scenario depicted in Figure 5, where two features  $f$  and  $g$  at version  $i$  and  $j$ , respectively, are very close to each other but topologically *Disjoint*. A geometric edit occurs that slightly modifies  $f_i$  into  $f_{i+1}$ , as depicted on the right side of the figure. Although the geometric change is very small, it

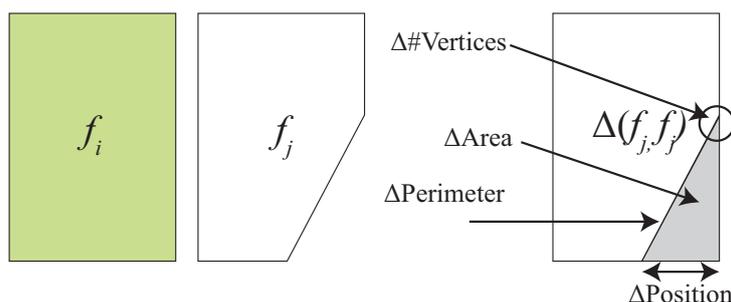
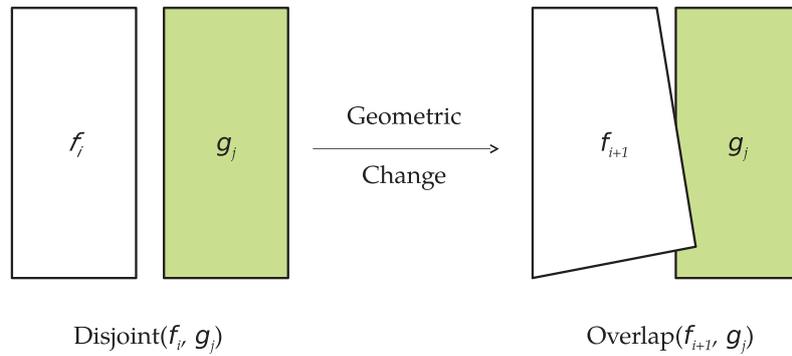


Figure 4. Geometric comparison between versions.



**Figure 5.** A small geometric change may correspond to a qualitative spatial change.

changes the topological relation holding between the two features. They are not *Disjoint* any longer but, rather, they *Overlap*, which, according to the theory of conceptual neighborhoods (Freksa 1991), is a sensible change. The conceptual neighborhood graph for topological relations (Egenhofer and Mark 1995) (as shown in Figure 6) indicates how topological relations may evolve following a continuous transformation. The distance on the graph evaluates the qualitative spatial difference between two topological configurations. For example, the relations *Disjoint* and *Overlap* are at distance two, because there is an intermediate configuration (*Meet*) between them.

### 3.5. Trustworthiness and reputation

Data trustworthiness ( $T$ ) and user reputation ( $R$ ) are discrete functions of time  $T, R: \{t_1, t_2, \dots, t_n\} \rightarrow [0, 1]$  defined on the set  $\{t_1, t_2, \dots, t_n\}$  of time instants when new feature versions are contributed that influence their value. Their values range in a continuum going from 0 (i.e. minimum reliability) to 1 (i.e. maximum reliability). To keep notation cleaner, in the following we avoid to make explicit the time dependency of trustworthiness and reputation: we write  $T(f_i)$  and  $R(a)$  instead of  $T(f_i, t)$  and  $R(a, t)$ , respectively.

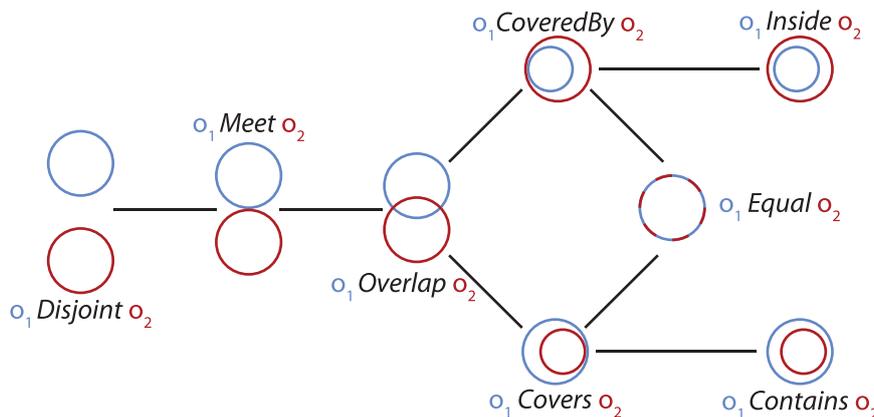
The reputation  $R(a)$  of an author  $a$  is the average of the trustworthiness values of all the feature versions authored by  $a$ :

$$R(a) = \frac{\sum_{f_i \in F(a)} T(f_i)}{|F(a)|} \quad (1)$$

where  $F(a)$  is the set of feature versions contributed by  $a$  until this point in time,  $|\cdot|$  denotes the cardinality of a set, and  $T(f_i)$  is the trustworthiness currently associated to the  $i$ -th of such feature versions.

All the complexity of the model lies in the definition of trustworthiness as an implementation of the many eyes principle (Raymond 1999): the more authors contribute to a feature (i.e. more versions), the higher the probability that errors are spotted and corrected – so that the contributed feature fits well to reality. Such a concept is captured by the notion of confirmation (resp. contradiction): the more versions conform (resp. differ) to each other with respect to the aspects described in Section 3.4, the higher (resp. lower) the trust of these versions.

We differentiate among three different types of confirmations, each capturing a different effect on the overall trustworthiness score: direct effect ( $T_{dir}$ ), indirect effect ( $T_{ind}$ ), and temporal effect ( $T_{tmp}$ ). Trustworthiness is defined as their weighted sum:



**Figure 6.** Conceptual neighborhood graph of topological relations, as defined in 9-Intersection model.

$$T(f_i) = w_{dir} \cdot T_{dir}(f_i) + w_{ind} \cdot T_{ind}(f_i) + w_{tmp} \cdot T_{tmp}(f_i) \quad (2)$$

where  $w_{dir}$ ,  $w_{ind}$ , and  $w_{tmp}$  are weights balancing the respective components. Since  $T \in [0, 1]$ , we have  $w_{dir} + w_{ind} + w_{tmp} = 1$ , with  $T_{dir}, T_{ind}, T_{tmp} \in [0, 1]$ .

### 3.5.1. Direct effect

The direct effect ( $T_{dir}$ ) expresses the similarity among all versions in the lineage of a feature. We call this effect “direct” as the information conveyed in the system by each version of a feature affects the trustworthiness of all versions of the same feature. So, this is a direct confirmation (or contradiction) of the state of a feature. Direct trustworthiness of a feature version increases as the attribute-wise difference to the average feature version diminishes, where the average feature version is obtained by averaging (attribute-wise) all versions in the lineage of a feature.

According to the tripartite classification of the aspects of an edit presented in Section 3.4, direct effect is regarded as consisting of three components: direct thematic effect ( $T_{dir,them}$ ), direct geometric effect ( $T_{dir,geom}$ ), and direct qualitative spatial effect ( $T_{dir,qual}$ ). The overall direct effect is obtained as the weighted sum of the three components:

$$T_{dir}(f_i) = w_{dir,them} \cdot T_{dir,them}(f_i) + w_{dir,geom} \cdot T_{dir,geom}(f_i) + w_{dir,qual} \cdot T_{dir,qual}(f_i) \quad (3)$$

where  $w_{dir,them}$ ,  $w_{dir,geom}$  and  $w_{dir,qual}$  are weights balancing the respective components. Since  $T_{dir} \in [0, 1]$ , we have  $w_{dir,them} + w_{dir,geom} + w_{dir,qual} = 1$ , with  $T_{dir,them}, T_{dir,geom}, T_{dir,qual} \in [0, 1]$ .

### 3.5.2. Indirect effect

The indirect effect ( $T_{ind}$ ) is designed to account for the effect of indirect confirmations on the trustworthiness of a feature. The concept of indirect confirmation was first introduced in Keßler and De Groot (2013). It is based on the consideration that if an author contributes a new version of a feature but leaves untouched nearby ones, then the latter are

likely to be correctly mapped. Accordingly, their trustworthiness must increase. Rather than considering all nearby features, it would probably be more appropriate to consider nearby landmarks, or more generally, outstanding features in the local context. There is a large literature addressing the challenge of defining what a landmark is – see Richter and Winter (2014) for an extensive survey – but this falls outside the scope of this article. So, in this work we consider indirect confirmation based on spatial proximity and temporal co-occurrence of feature versions. That is, for a feature version  $f_i$  to be indirectly confirmed by another feature version  $g_j$ , two conditions must be fulfilled:

- (a) (temporal co-occurrence)  $f_i$  exists when  $g_j$  is created;
- (b) (spatial proximity)  $g_j$  is within a given range from  $f_i$ .

The example in Figure 7 illustrates the indirect effect in case of versions  $f_i$ ,  $a_j$ , and  $b_k$  of three different features. Figure 7(a) represents the lifetime of the feature versions and can be referred to check temporal co-occurrence, while Figure 7(b) depicts spatial proximity, with the dotted line denoting the proximity range of feature  $f$ . The entities that satisfy the temporal co-occurrence or spatial proximity are depicted in green. Only the trustworthiness of those entities that satisfy both criteria (green on both sides of the figure) are indirectly affected by  $f_i$ ; those in gray are not affected. Version  $f_i$  is created during  $a_j$  and  $b_k$  lifetime, with  $b_k$  also falling in the proximity area. Both versions  $a_j$  and  $b_k$  satisfy the condition on temporal co-occurrence, but only  $b_k$  satisfies the condition on spatial proximity. So,  $b_k$  is indirectly confirmed by  $f_i$ , while  $a_j$  is not. Note that the feature version  $a_{j+1}$  is not affected by  $f_i$  as it is still not existing. Conversely, when  $a_{j+1}$  will start to exist, it will influence the trustworthiness of  $f_i$  – assuming that the condition on spatial proximity will also be satisfied.

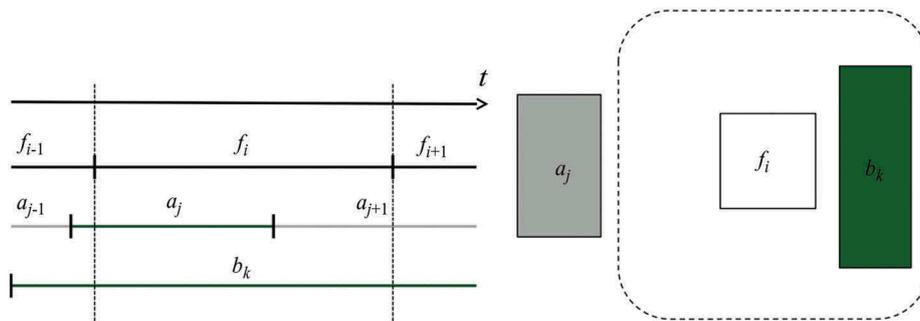


Figure 7. Versions that determine indirect effect.

Similarly to the direct effect, also the indirect effect consists of three components: indirect thematic effect  $T_{ind,them}$ , indirect geometric effect  $T_{ind,geom}$ , and indirect qualitative spatial effect  $T_{ind,qual}$ . The overall indirect effect is obtained as the weighted sum of the three components:

$$T_{ind}(f_i) = w_{ind,them} \cdot T_{ind,them}(f_i) + w_{ind,geom} \cdot T_{ind,geom}(f_i) + w_{ind,qual} \cdot T_{ind,qual}(f_i) \quad (4)$$

where  $w_{ind,them}$ ,  $w_{ind,geom}$ , and  $w_{ind,qual}$  are weights balancing the respective components. Since  $T_{ind} \in [0, 1]$ , we have  $w_{ind,them} + w_{ind,geom} + w_{ind,qual} = 1$ , with  $T_{ind,them}, T_{ind,geom}, T_{ind,qual} \in [0, 1]$ .

### 3.5.3. Temporal effect

The temporal effect ( $T_{tmp}$ ) is designed to account for temporal confirmation, that is, the longer a version persists in the system, the higher the probability that it is well mapped. Accordingly, we want that the trustworthiness of a feature version increases over time as it remains unaltered, converging to  $T = 1$  as  $t$  tends to infinity. Again, this is a derivation of the many eyes principle: a version with serious and remarkable errors is more likely to be modified sooner than a version with negligible errors.

The temporal effect determines the increment of trustworthiness until the maximum possible level. As illustrated in Figure 8, the temporal effect asymptotically increases toward 1, as described by the function:

$$T_{tmp}(f_i) = \frac{t(f_i)}{t(f) + c} \quad (5)$$

where  $t(f_i)$  is the lifetime of feature version  $f_i$ ,  $t(f)$  is the overall lifetime of feature  $f$  (i.e. the sum of the lifetimes of all versions of  $f$ ), and  $c$  is a parameter that can be adjusted to modify the curve slope.

## 3.6. A possible implementation of direct and indirect trustworthiness

In previous sections, we introduced our model for trustworthiness and reputation and gradually broke down trustworthiness into finer pieces. The finest level comprises thematic, geometric, and qualitative

spatial aspects for both direct and indirect effects; specifically, direct thematic effect ( $T_{dir,them}$ ), direct geometric effect ( $T_{dir,geom}$ ), direct qualitative spatial effect ( $T_{dir,qual}$ ), indirect thematic effect ( $T_{ind,them}$ ), indirect geometric effect ( $T_{ind,geom}$ ), and indirect qualitative spatial effect ( $T_{ind,qual}$ ).

In this section, we propose one possible implementation of these fine-grained components of trustworthiness. The empirical evaluation of the model presented in Section 5 is based on this implementation. We leave for future work the task of deriving more complex implementations.

### 3.6.1. Direct thematic effect

The direct thematic effect ( $T_{dir,them}$ ) considers how thematic attributes vary among different versions. If a set of attributes is used more consistently throughout different versions of a feature, then the contribution of these attributes to the direct thematic effect is higher.

A possible behavior can be expressed as follows:

$$T_{dir,them}(f_i) = 1 - \frac{noDiffThem(f_i)}{n} \quad (6)$$

where  $f_i$  is the version under assessment,  $n$  is the number of versions of the feature  $f$ , and  $noDiffThem(f_i)$  is the number of versions of  $f$  that have a different thematic attribute set than  $f_i$ . As mentioned before, more complex behaviors would take into consideration other metrics for the attributes, for example, not only by counting the number of versions with different attribute sets, but also measuring the lexical distance among them.

### 3.6.2. Direct geometric effect

The direct geometric effect ( $T_{dir,geom}$ ) addresses the difference between a feature version  $f_i$  and the average feature version  $f_{avg}$  in terms of geometric properties. Let

$$\Delta p = 1 - \frac{|p_i - p_{avg}|}{|p_i - p_{avg}| + c} \quad (7)$$

be a measure of the similarity between  $f_i$  and  $f_{avg}$  for what concerns the generic geometric property  $p$ ,

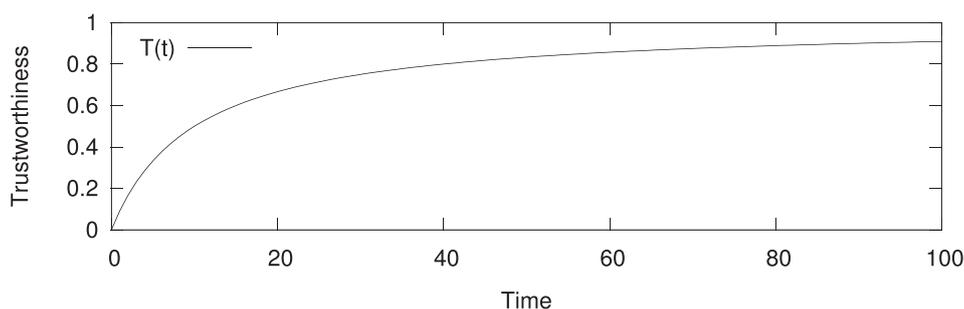


Figure 8. Temporal effect contribution on trustworthiness.

where  $p_i$  is the value of the property for  $f_i$ ,  $p_{avg}$  is the average property value over the versions currently present in the feature version lineage, and  $c$  is a parameter that can be adjusted to modify the curve slope. Then, the direct geometric effect can be expressed as

$$T_{dir,geom} = \frac{\sum_{p \in P} \Delta p}{|P|} \quad (8)$$

where  $P = \{Area, Perimeter, VertexNumber, VertexPosition\}$  is the set of considered geometric properties and  $|P|$  denotes its cardinality. Note that the set of considered properties can be altered according to those available in the VGI at hand.

### 3.6.3. Direct qualitative spatial effect

The direct qualitative spatial effect ( $T_{dir,qual}$ ) considers the change in qualitative relations among different versions. For this effect, speaking of average values makes no sense, as in general qualitative relations range in a discrete set with no total order defined. Therefore, we count the number of times a specific qualitative spatial relation occurs between the feature version at hand  $f_i$  and the other features in the system. To reduce the computational complexity, we suggest to filter the features to be considered. A simple filter can be achieved by considering only those features falling within a given distance  $d$  from the feature version  $f_i$ . Alternatively, filtering can be based on more sophisticated criteria that only select features satisfying given thematic or geometric constraints. For example, one may select only buildings whose area is bigger than a given threshold.

For simplicity, let us consider only current neighbors of feature version  $f_i$ , i.e. the set of feature versions  $N$  that are alive when version  $f_i$  is generated and whose distance from  $f_i$  is not greater than  $d$ . We denote by  $F = \{f_1, \dots, f_i\}$  the set of all versions of feature  $f$ , with  $f_i$  being its latest version. Let

$$rel_{j,k} = rel(f_j, k) \text{ with } j \in [1, i], k \in N \quad (9)$$

be the qualitative spatial relation holding between the  $j$ -th version of  $f$  and the  $k$ -th neighbor of  $f_i$ . Finally, we use Iverson bracket notation and define

$$C(rel_{i,k}) = \sum_{j=1}^i [rel_{j,k} = rel_{i,k}] \quad (10)$$

as the number of times the relation holding between  $f_i$  and the generic neighbor  $k$  also holds for previous versions of  $f$ .

Then, we can define direct qualitative spatial trustworthiness as

$$T_{dir,qual}(f_i) = \frac{1}{|N|} \cdot \sum_{k \in N} \frac{C(rel_{i,k})}{n} \quad (11)$$

where  $n$  denotes the total number of versions of feature  $f$ , and  $|N|$  denotes the number of neighbors of  $f_i$ . Said differently, we first count for each neighbor  $k \in N$  of  $f_i$  the number of times the relation  $rel_{i,k}$  also holds for past versions  $f_j$  (with  $j < i$ ) of  $f$  and normalize it by dividing by the number of versions of  $f$ . Then, we average these values over all neighbors of  $f_i$ .

Note that, although previous versions of neighbors are not considered here, they will affect the trustworthiness of  $f_i$  indirectly, provided that we chose a commutative filtering function to generate set  $N$  (such as distance).

### 3.6.4. Indirect effects

One very easy approach to model indirect effects ( $T_{ind,them}$ ,  $T_{ind,geom}$ ,  $T_{ind,qual}$ ) is to consider the reputation of the author  $a$  that contributes the feature version  $g_j$  triggering an indirect confirmation for feature  $f_i$ .

We defined the overall reputation of an author  $a$  in Equation 1 as the average trustworthiness of the feature versions authored by  $a$ . Trustworthiness, as defined in Equation 2, consists of direct, indirect, and temporal effects. In turn, direct (Equation 3) and indirect (Equation 4) effects consist of thematic, geometric, and qualitative spatial aspects. By substituting Equations 3 and 4 in Equation 2 and by applying basic algebraic transformations, we can refactor the overall trustworthiness in terms of aspects (rather than by effects):

$$T(f_i) = w_{them} \cdot T_{them}(f_i) + w_{geom} \cdot T_{geom}(f_i) + w_{qual} \cdot T_{qual}(f_i) + w_{tmp} \cdot T_{tmp}(f_i) \quad (12)$$

with the generic component  $w_x \cdot T_x(f_i)$ ,  $x \in \{them, geom, qual\}$  equal to

$$w_x \cdot T_x(f_i) = w_{dir} \cdot w_{dir,x} \cdot T_{dir,x}(f_i) + w_{ind} \cdot w_{ind,x} \cdot T_{ind,x}(f_i) \quad (13)$$

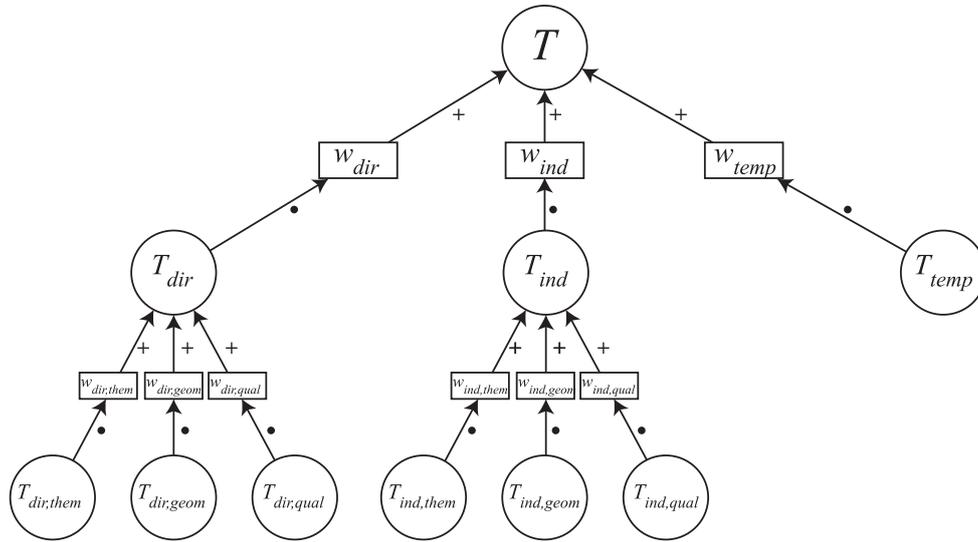
Equivalently, we can refactor the overall reputation of author  $a$  as follows:

$$R(a) = R_{them}(a) + R_{geom}(a) + R_{qual}(a) + \frac{\sum_{f_i \in F(a)} w_{tmp} \cdot T_{tmp}(f_i)}{|F(a)|} \quad (14)$$

with the generic component  $R_x(a)$ ,  $x \in \{them, geom, qual\}$  equal to

$$R_x(a) = \frac{\sum_{f_i \in F(a)} w_x \cdot T_x(f_i)}{|F(a)|} \quad (15)$$

where  $F(a)$  is the set of feature versions contributed by  $a$  until this point in time,  $|\cdot|$  denotes the cardinality of a set, and  $w_x \cdot T_x(f_i)$  is the trustworthiness



**Figure 9.** Graphic explanation of the composition of trustworthiness (reputation looks identical).

aspect (Equations 12 and 13) currently associated to the  $i$ -th of such feature versions.

Then, we can define indirect trustworthiness effects as

$$T_{ind,them}(f_i) = R_{them}(a), \quad (16)$$

$$T_{ind,geom}(f_i) = R_{geom}(a), \quad (17)$$

$$T_{ind,qual}(f_i) = R_{qual}(a) \quad (18)$$

Figure 9 presents a summary of how trustworthiness is composed.

#### 4. System architecture

In Section 3, we presented a model to derive trustworthiness and reputation scores for VGI. This section presents and analyzes a system architecture that implements the model and that we call TandR, shortly for Trustworthiness and Reputation.

##### 4.1. Requirements and features

The TandR framework uses feature evolution information. The outputs consist of a trustworthiness score for each spatial feature version and a reputation score for each contributor. The framework is able to manage only VGI systems that adhere to the model of Section 3; that is, feature evolution information is organized in versions, where each version is the result of an *atomic edit*, either a creation or a modification or a cancellation.

The framework must perform data analysis on multiple VGI system sources; it does not have to provide direct support to each specific system format, but it must be able to analyze data from different sources; such data have to be brought back to an

established format that is flexible and general enough to be able to manage all necessary information.

The problem of giving the right structure to feature evolution information can be solved by organizing data in versions and defining a common format for information sources. Provenance information was already adopted and properly structured in Keßler, Trame, and Kauppinen (2011), as already discussed in Section 2.2. Figure 1 shows the ontology designed in this work, which correlates spatial data taken from the OSM VGI system with provenance information. We use such an ontology as a starting point to introduce, in Section 4.2, a more general ontology that models VGI information.

The model introduced in Section 3 describes a basic strategy for scores calculation; nonetheless, it can handle changes and extensions. Extensibility of the framework is implemented through a mechanism of module selection, with which the framework can handle different implementations of the trustworthiness and reputation model: each module has the task of managing a different implementation. The selection of a given module is made by the user via a string passed to the framework configuration.

In addition to the last trustworthiness and reputation scores, the framework keeps record of historic scores that show how trustworthiness and reputation evolve over time. The model implementation follows events time line; therefore, computation takes into account a series of trustworthiness and reputation scores, each representing the level of “reliability” that versions and users had in the moment the scores calculation refers to.

##### 4.2. Data models

Data are modeled following two ontologies: HVGI and TandR. The historical VGI (HVGI) ontology,

shown in Figure 10, models the domain of VGI systems provenance information and is used to manage VGI feature versions. This ontology includes many concepts defined in external ontologies, notably, the OSM provenance information ontology proposed by Keßler, Trame, and Kauppinen (2011) that defines the concept of FeatureState (corresponding to our feature version). Also, it includes the ontology defined by the Open Geospatial Consortium (OGC) to manage feature geometry, with the Geometry class and its subtypes Point, LineString, and Polygon. Finally, it includes the Friend Of A Friend (FOAF) ontology to manage user details.

The TandR ontology (as shown in Figure 11) models trustworthiness and reputation scores domain. It allows us to represent the score data of the model introduced in Section 3. The ontology provides definition of Trustworthiness and Reputation concepts, linking them to the concept of Effect. This can consist of one or more Aspects, as for the case of direct and indirect effects that are broken down into semantic, geometric, and qualitative aspects. The concepts of Effect and Aspect provide flexibility to the model, in the sense that they are general and can be used to extend or modify the model proposed in this work. Note that Trustworthiness and Reputation are connected to FeatureState and User concepts, respectively, both defined in the OSM provenance information ontology (Keßler, Trame, and Kauppinen 2011). TrustworthinessValue and ReputationValue group two bits of information: the numerical score value and a timestamp, indicating the point in time since when the score is valid. Thanks to this information, it is possible to reconstruct the evolution of the trustworthiness of a feature version and of the reputation of a user over time. The Effect concept is related to two more concepts. The EffectDescription provides the name of the effect, a human-readable description of it, and relates to the person that defines it. The EffectValue reports the numeric value of the effect and the date from which that value is valid. The Aspect concept is specular to the Effect.

### 4.3. Framework overview

In this section, we describe the system components diagram of the TandR framework (as shown in Figure 12). The system uses semantic technologies, hence information persistence is managed by the Parliament triple store (<http://parliament.semwebcentral.org/>). Parliament uses the quadruple pattern (*graph {subject predicate object}*) and is spatially enabled via the GeoSPARQL standard (<http://www.opengeospatial.org/standards/geosparql/>). The system adopts the Parliament libraries available in the Java language, that provide an API to interact with the underlying triple store. As shown in Figure 12, the triple store manages three different components, each of

which represents a graph that contains a set of related RDF (<https://www.w3.org/RDF/>) triples: the graph that contains VGI data, the graph that contains trustworthiness and reputation scores, and the graph that contains results of scores validation.

The TandR framework is implemented in Java. To manage interactions with the triple store and process data, TandR uses two external components: the Java Topology Suite (JTS) (<https://github.com/locationtech/jts/>) and Jena (<https://jena.apache.org/>). The former provides support for spatial operations, while the latter collaborates with Parliament to manage the connection to the triple store.

Data sources are divided into two kinds. The first one, the most commonly used, needs a layer that adapts data from the VGI system format to the application format, as specified by the HVGI ontology. The second kind of source provides data directly to the format required by the application. OSM represents an example of the first kind of source that requires the development of a component that translates the OSM history file into the application format: we called such a component OSH2RDF. In Figure 12, the components in green indicate that they have been implemented in our prototype, while the components in gray indicate that they can be added in future developments.

The framework supervises the following activities:

- (1) **Installation.** The installation process aims at preparing a triple store to host data. More in detail, the installation process carries out the management and creation of graphs, the VGI system data import, and spatiotemporal indexing.
- (2) **Scores Computation.** As already mentioned, multiple scores are provided for versions and users, each associated with a validity date. The latest score available is always the most reliable, due to the fact that it was computed with a greater amount of provenance information. The obtained scores are structured and made persistent in such a way that they are properly differentiated by version date.
- (3) **Scores Validation** Scores validation. compares the obtained scores against a ground truth data-set, in order to assess their quality. We will explore how validation works in Section 5.

In the framework configuration, it is possible to select which of the three activity the user wishes to perform or even more than one activity at a time.

### 4.4. Framework behavior

Events and edits are considered in temporal order. Therefore, the framework will take into account the order in which versions were generated and process



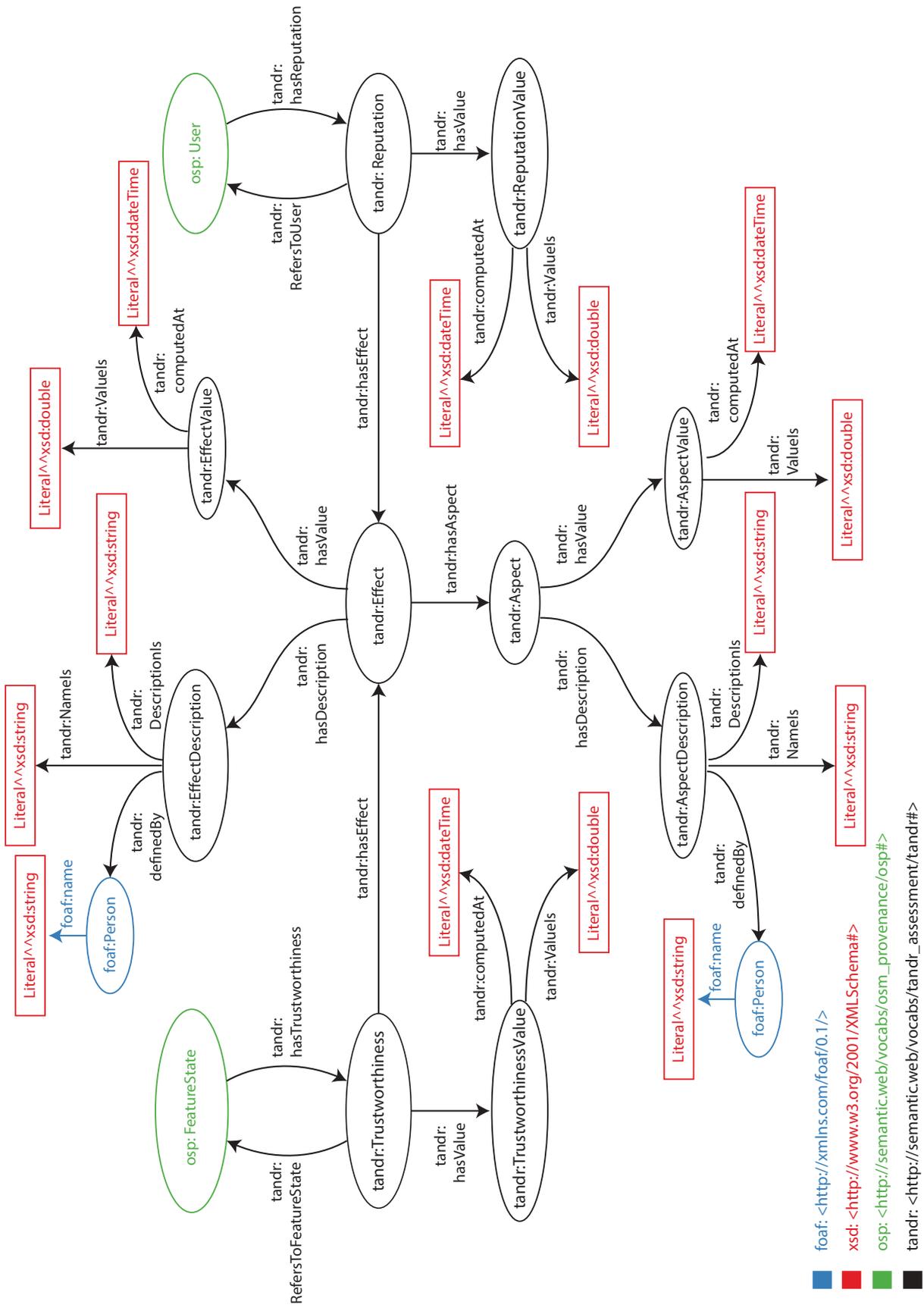


Figure 11. Domain ontology: trustworthiness and reputation.

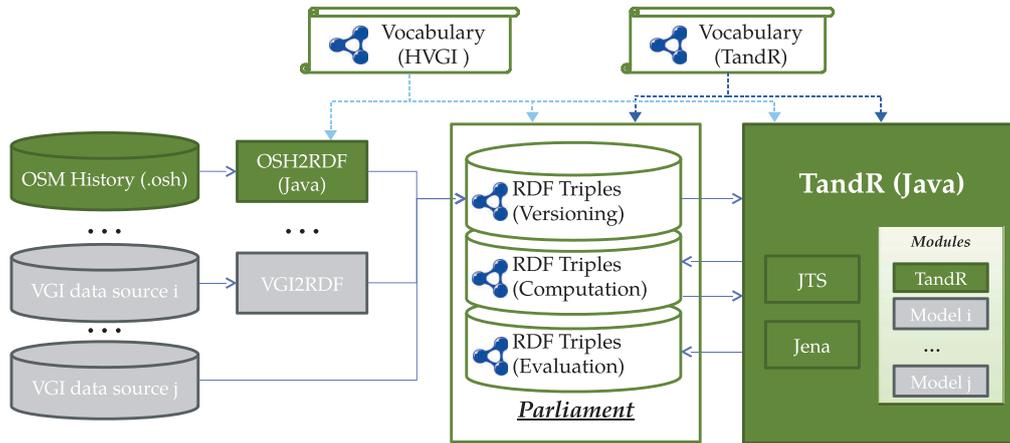


Figure 12. The TandR framework: component diagram.

them accordingly. One way to achieve this result is presented in the following algorithm:

Retrieve all dates in which at least one feature version was created.

For each date:

Retrieve all versions whose lifetime begins in that date.

For each version:

Delegate scores calculation to selected module.

Reputation and trustworthiness scores may be calculated according to our model (Section 3.6). After the first time a version is processed, trustworthiness is updated whenever a new feature version (direct effect) or a nearby version (indirect effect) is created; the temporal effect is updated every time either of the other effects are altered. See the following algorithm:

Retrieve, if any, previous versions to current version  $f_v$ .

Retrieve, if any, living  $f_v$  neighbors.

Trigger Trustworthiness calculation of  $f_v$ .

Trigger Reputation calculation for  $f_v$  author.

For each  $f_u$  previous version of  $f_v$ :

Update Trustworthiness of  $f_u$ .

Update Reputation of  $f_u$  author.

For each living nearby version  $n_v$ :

Update Trustworthiness of  $n_v$ .

Update Reputation of  $n_v$  author.

## 5. Empirical validation

In order to prove the validity of the proposed model, we should show that feature versions with higher trustworthiness resemble more closely the features

mapped in authoritative data-sets. In this section, we provide a qualitative validation of the model by (1) running our TandR Java framework on VGI data to derive trustworthiness and reputation scores and (2) selecting a feature  $f$  and comparing its lowest  $f_{min}$ , medium  $f_{med}$ , and highest  $f_{max}$  trustworthy versions against the corresponding feature  $f_t$  from a ground truth data-set. The model proves valid if the similarity  $\sigma$  of the three feature versions to the reference feature grows as trustworthiness grows. That is, the following system of equations should be satisfied:

$$\begin{cases} \sigma(f_{min}, f_t) \leq \sigma(f_{med}, f_t) \leq \sigma(f_{max}, f_t) \\ T(f_{min}) \leq T(f_{med}) \leq T(f_{max}) \end{cases} \quad (19)$$

where the similarity function  $\sigma$  accounts for the same aspects used by the TandR framework to compute trustworthiness and reputation scores. Note that the model does not require a comparison to a ground truth data-set to work. This is only done to prove the validity of the model.

### 5.1. Data sources

As VGI data source we use the OSM full history dump file (<http://planet.osm.org/planet/full-history/>), which contains the evolution history from the beginning of the project until today. The ground truth data-set was obtained from Open Government Wien (<https://open.wien.gv.at/>), which provides Austrian land cover data; a view of such government data is shown in Figure 13 and is about the municipality of Vienna in 2012.

For the experiment reported in this article, we restrict ourselves to the spatial window depicted in Figure 14 and to a temporal window of four years starting January 2010 and ending December 2013. The resulting data-set, obtained from the OSM full history dump, consists of 834 feature versions belonging to 552 different geographic features edited by 38 users.

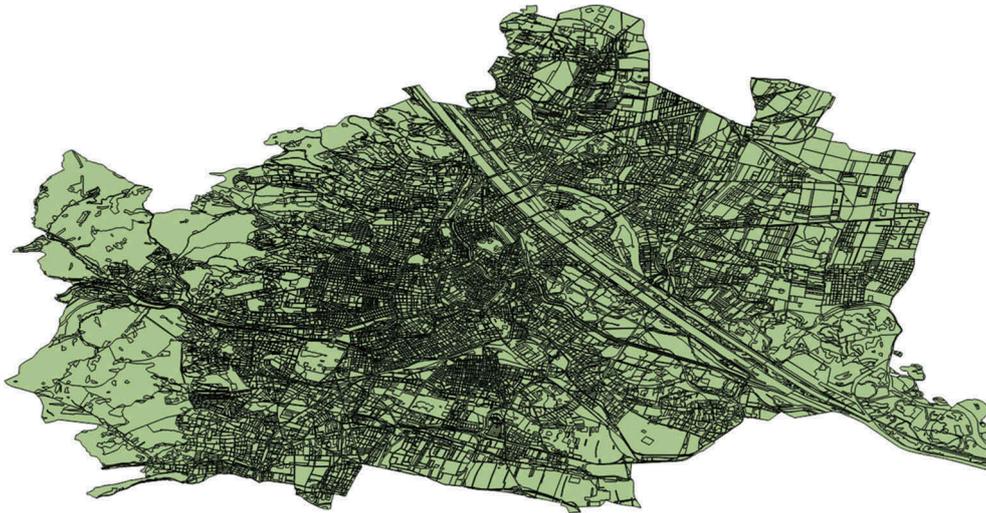


Figure 13. Wien ground truth data-set from open government Wien.

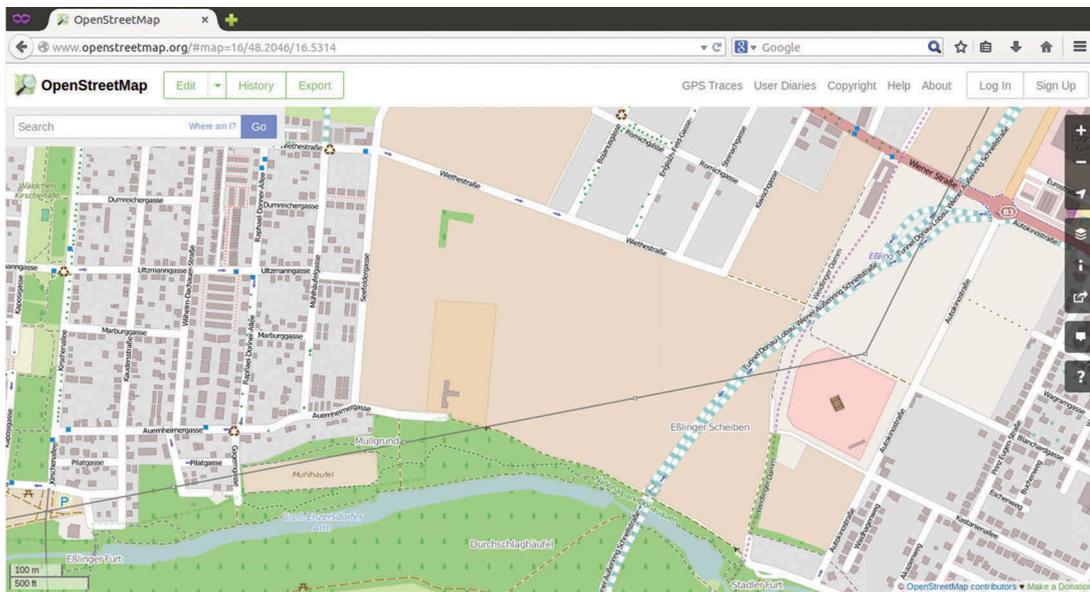


Figure 14. Screenshot of the selected area in OSM.

## 5.2. Experiment setup and validation methodology

First, we run the TandR framework on the data-set defined above with the following weights:  $w_{dir} = w_{ind} = w_{tmp} = \frac{1}{3}$ ,  $w_{dir,them} = w_{dir,geom} = w_{dir,qual} = \frac{1}{3}$ ,  $w_{ind,them} = w_{ind,geom} = w_{ind,qual} = \frac{1}{3}$ . For the slope parameters of the temporal effect (Equation 5) and the direct geometric aspect (Equation 7), we set the values 500 and 1, respectively. For the geometric aspect we adopted the measures  $\{Area, Perimeter, VertexNumber\}$  and for the qualitative aspect we considered topological relations as defined in the 9-Intersection model (Egenhofer 1989). For the thematic aspect we used the formula in Equation 6.

As a second step, we performed a qualitative validation. The validation methodology associates to each ground truth feature  $f_t$  the corresponding VGI feature  $f$  and compares each version  $f_v$  of  $f$  against the

corresponding feature  $f_t$  from the ground truth data-set. In other words, we consider the feature  $f_t$  from the ground truth as if it was an extra version of the corresponding VGI feature. Then, the comparison is done by applying the same calculations used for computing the direct trustworthiness values only (since the ground truth data-set does not evolve over time, it does not make sense to account for the temporal and the indirect effects).

Since the ground truth data-set does not report thematic information, the comparison is performed only considering the direct geometric and direct qualitative aspects. Accordingly, the first-level weights of the model (Equation 2) have been set as follows:  $w_{dir} = 1$ ,  $w_{ind} = w_{tmp} = 0$ . The finer-grained weights for the direct effect (Equation 3) have been set to:  $w_{dir,geom} = w_{dir,qual} = 0.5$ ,  $w_{dir,them} = 0$ .

The high-level algorithm used to validate trustworthiness indexes can be described as follows:  
For each feature  $f_i$  in the ground truth data-set:

Retrieve corresponding VGI feature  $f$ .

For each version  $f_v$  of the retrieved VGI feature:

Compare the geometry of  $f_v$  with the geometry of  $f_i$ .

Apply comparison logic to obtain the similarity score  $\sigma(f_v, f_i)$ .

To identify correspondences between the VGI and the ground truth data-set, we consider the versions  $f_v$  of VGI feature  $f$ , whose geometry intersects  $f_i$ . Hence, for each feature  $f_i$  in the ground truth data-set, we retrieve the VGI feature  $f$  that has in its lineage the version  $f_v$  with the largest intersection area with  $f_i$ . The similarity score  $\sigma(f_v, f_i)$  is obtained by comparing the geometries of the VGI features against those of the authoritative ones, as explained in Section 3.6.

### 5.3. Results

To showcase the results obtained in our experiment, we report in Figure 15 the computed trustworthiness values for the OSM feature 32506491 over time. Note that while edits are identified by the date and time they occurred, the x-axis is not scaled over time. That is, dates are nominal values in this plot. The different versions of the feature are reported in different colors. Also, note that, although this specific feature (in the time range considered in our experiment) consists of 16 versions, the number of plotted trustworthiness values are much more. The first

data point of each version (color) is equal to the reputation of the version's author at the time of insertion in the VGI system. All the other data points in a feature version life correspond to an adjustment of the trustworthiness ignited by indirect confirmations.

In general, we can observe how the trustworthiness value swings quite smoothly up and down according to the impact of the effects and aspects discussed in previous sections. However, at the beginning of the time interval the trustworthiness value jumps quite abruptly from 0 to approximately 0.3. This might be related to the specific choice of the slope parameters influencing the behavior of the temporal aspect (Equation 5) and of the direct geometric aspect (Equation 7). The value of the parameters shall be tuned in future experiments to obtain a smoother effect.

As an example, we show in Tables 1 and 2 trustworthiness and similarity scores obtained for two polygons: OSM feature ids 38838966 and 45275690, respectively. The trustworthiness scores are computed using all the effects and aspects discussed in Section 3.6. The similarity values are computed using only the direct geometrical aspect and the direct qualitative spatial aspect, as explained in the previous section. Both types of values, trustworthiness and similarity, do not have an absolute meaning. Rather, they have to be interpreted in a relative manner among several versions of the same features. So, for example, we can read that version 4.0 of feature 38838966 is the most trustworthy among all of its versions, but we cannot infer anything about how reliable this version is with respect to other features.

Both the examples reported below satisfy the qualitative validation inequalities reported in Equation

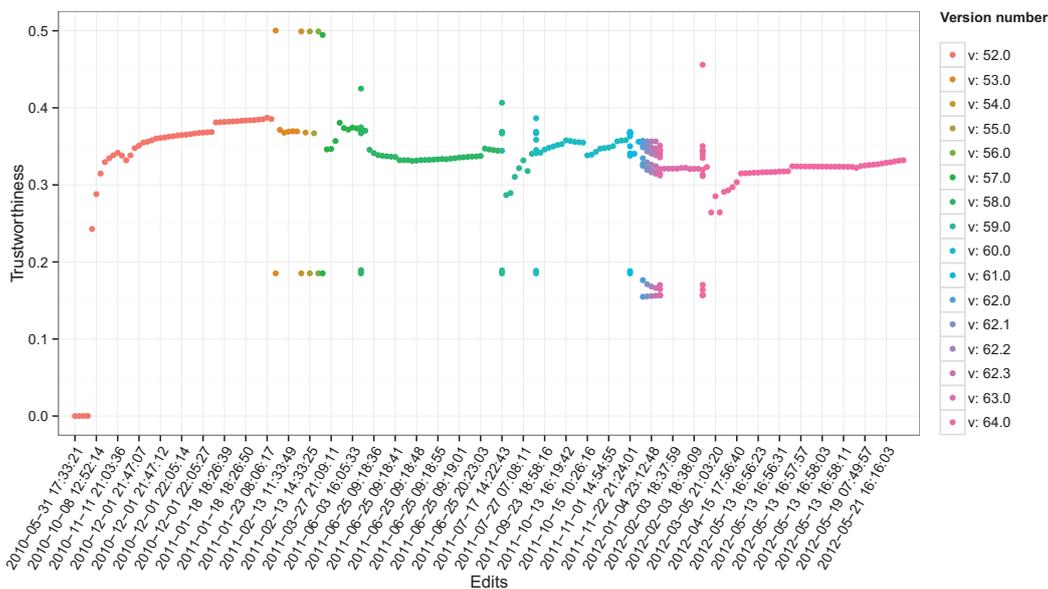


Figure 15. Trustworthiness values for the OSM feature 32506491 varying over edits and feature versions.

**Table 1.** Validation indexes for feature id 38838966.

Version Number	Trustworthiness	Similarity
3.0	0.1179	0.4166
4.0	0.6138	0.3888
5.0	0.4462	0.3888
6.0	0.4874	0.3888

**Table 2.** Validation indexes for feature id 45275690.

Version Number	Trustworthiness	Similarity
3.0	0.5065	0.3749
3.1	0.4419	0.3749

19. For feature 38838966 (Table 1), the qualitative validation condition reported in Equation 19 is satisfied: it is enough to set  $f_{min}$ ,  $f_{med}$ , and  $f_{max}$  to versions 3.0, 5.0 (version 6.0 would work as well), and 4.0, respectively. Note that we have several versions (namely 4.0, 5.0, and 6.0) of feature 38838966 with different trustworthiness but same similarity scores. This is due to the fact that we could only apply the geometric and qualitative spatial aspects to compare against the ground truth data-set, while these versions are generated by thematic edits.

For the other sample feature 45275690 (Table 2), we only have two versions. In this case, one can simply let two of the versions in Equation 19 be the same. For example, we can set  $f_{min} = f_{med}$  and  $f_{max}$  to versions 3.1 and 3.0, respectively.

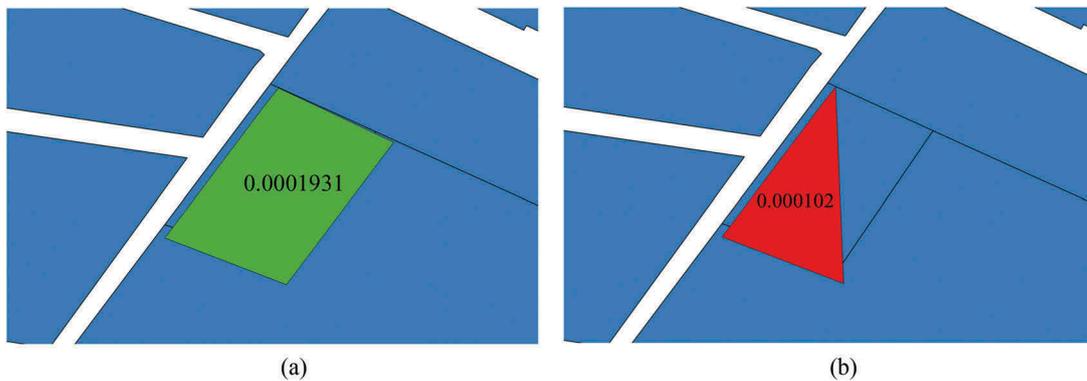
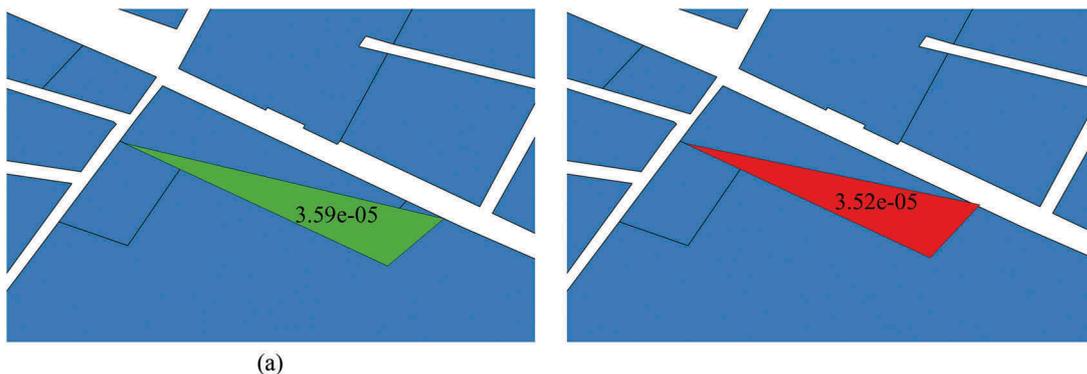
Figures 16 and 17 allow for a visual comparison of the versions of the selected VGI features with minimum

(in red) and maximum (in green) trustworthiness against the ground truth data-set (in blue) in the background. Figure 16 reports about VGI feature 38838966 and clearly shows how better feature version 4.0 (Figure 16(a)) geometrically matches the reference feature from the ground truth data-set. The latter is visible in Figure 16(b), underneath version 3.0 of the VGI feature (in red). Figure 17 reports about VGI feature 45275690: Figure 17(a) shows version 3.0 with maximum similarity to the ground truth data-set; Figure 17 (b) shows version 3.1 with minimum similarity. In this case, neither of the two versions really fits well to the ground truth reference, yet, according to the similarity measures used, version 3.1 results the best fit. In this case the lineage of the feature (with only two versions) is definitely too short to efficiently implement the many eyes principle.

All results in the selected territory portion seems to indicate that our model of trustworthiness provides a valid measure of VGI data quality with respect to ground truth data. Collected data encourages this approach and shows its validity; nevertheless, it is necessary to improve it and perform more exhaustive tests.

## 6. Conclusions and future work

VGI systems collect spatial information from Internet users and follow the principles that stand behind crowd-source information. The scope of map-based VGI

**Figure 16.** Two versions of feature 38838966 superimposed on the ground truth data-set.**Figure 17.** Two versions of feature 45275690 superimposed on the ground truth data-set.

systems, such as OSM, is to build a global map by filling spatial information and to maintain it freely accessible. One key factor for developing and popularizing these systems are “web users” and the contribution they provide; the higher is the number of contributors, the larger is the collected information, in terms of both amount and quality. The definition of a method that quantifies how good VGI is constitutes an active research topic. Among proposed methods, we based our work on the approach taken by Keßler, Trame, and Kauppinen (2011) and Keßler and De Groot (2013).

We presented a model to derive trustworthiness and reputation scores for geographic features and contributors of a VGI system, respectively. The model relies on provenance information and applies to all VGI systems that support versioning, where each version is the outcome of an atomic edit action, either creation, modification, or deletion. The model provides a trustworthiness score for each feature version and a reputation score for each author.

The model was implemented as a Java application using semantic technologies: ontologies have been developed for the representation of data that was made persistent with triplestores. The system architecture is thought of as a framework that can be easily extended and modified by introducing new modules that implement different score calculation models.

The software framework was run on an OSM data-set (full history dump). The obtained scores were validated against a ground truth data-set provided by an authoritative source. For each authoritative feature, we selected the corresponding VGI feature and verified that versions with increasing trustworthiness scores have increasing similarity to the reference feature. In the validation, we only applied direct geometric and qualitative spatial aspects, but we did not consider the thematic aspect and the temporal effect. The results met the expectations since an increase in trustworthiness corresponds to an increase in the similarity to the authoritative reference feature.

The current implementation of the framework can be extended in many directions. We foresee to extend the model by also considering the typical region of contribution of single users. That is, typically the same user contributes information about the same geographic region(s). Consequently, it is reasonable to derive a reputation score for this user that is a proxy measure of her/his local knowledge. Arguably, over time the user reputation will increase and, accordingly, the trustworthiness of the features authored by her/him. Yet, if the same user starts contributing information about a different area (maybe because she/he moved to a different city or state), there is no reason to assume that she/he has good local knowledge about the new region. Thus, the reliability of her/his contributions should be rated accordingly. This is especially true for the thematic aspect of edits but not much for

her mapping and cognitive skills (i.e. the geometric and qualitative aspect, respectively).

The reputation of a user should also account for the frequency of contributions, that is, if a user becomes inactive, her/his reputation should diminish over time. This behavior might be modeled similarly to the temporal effect of trustworthiness. In this case, the curve should be monotonically decreasing and converging to 0.

The computation of trustworthiness scores might be improved by introducing a temporal window. Indeed, as of now, trustworthiness scores are updated by comparing each version in a feature lineage against the “average feature version”. This might lead to very strange behaviors in case a feature disappears or changes drastically in the real world. For example, consider the case when a building is demolished. In this case, the older versions of the feature will be completely different than the newer versions. With time, the trustworthiness of the newer versions will increase while the trustworthiness of the older versions will decrease. The reputation of their authors will be altered accordingly. Yet, the deterioration of the reputation of the authors of old versions should be mitigated as at the time of contribution the building was still existing. To solve this issue, we suggest to introduce a temporal window that, at the moment of updating trustworthiness scores, only selects a subset of the versions in a feature lineage, leaving untouched very old versions and, consequently, the reputation of their authors.

Finally, the framework can be more thoroughly validated by considering other case studies, the thematic aspect, more domain ontologies, and a wider range of qualitative spatial relations (such as directional and distance operators).

## Notes on contributors

*Paolo Fogliaroni* is a post-doctoral researcher in the Department for Geodesy and Geoinformation, Vienna University of Technology, Austria. He received his BS and MS degrees from the University of L'Aquila in 2005 and 2008, respectively. He received his Doctoral degree (Dr.-Ing.) from University of Bremen, Germany, in 2012. His research interests are in the areas of spatial Human-Computer Interaction (HCI), spatial databases, spatial cognition, and Qualitative Spatial Representation and Reasoning (QSR). His latest research endeavors are in the areas of Spatial HCI with cutting-edge media technology, Spatial and Geographic Information Science, and Virtual Spatial Cognition. He investigates how the virtual augmentation of physical reality affects spatial cognition and what are the best metaphors to spatially interact with a virtually augmented world.

*Fausto D'Antonio* obtained his BS and MS degrees from the University of L'Aquila in 2011 and 2014, respectively. He was on Erasmus exchange with Vienna University of Technology. He discussed an MS thesis on evaluation of data trustworthiness through the analysis of historical data within VGI. His research interests are VGI and Semantic

Web. He is currently collaborating with University of L'Aquila and working in a private company Thales Alenia Space on On-Board Software for Space and Avionic Systems.

**Eliseo Clementini** is an associate professor of computer science at the Department of Industrial and Information Engineering and Economics of the University of L'Aquila (Italy). He received his M.Eng. in Electronics Engineering from University of L'Aquila in 1990 and his Ph.D. in Computer Science from University of Lyon (France) in 2009. He has been a visiting professor at the National Center for Geographic Information and Analysis of the University of Maine, at the Department of Geography and Geomatics of the University of Glasgow, at the Department of Geography of the University of Liege, and at the Department of Geodesy and Geoinformation of Vienna University of Technology. His research interests are mainly in the fields of spatial databases and geographical information science. One of his well-known contributions is the DE+9IM model for topological relations, which is part of Open Geospatial Consortium recommendations and ISO/TC 211 standard. He is a member of the program committee of important conferences such as the Conference on Spatial Information Theory (COSIT), the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), and the International Conference on Geographic Information Science (GIScience).

## ORCID

Paolo Fogliaroni  <http://orcid.org/0000-0002-0578-8904>  
 Eliseo Clementini  <http://orcid.org/0000-0002-3057-7105>

## References

- Antoniou, V., and A. Skopeliti. 2015. "Measures and Indicators of VGI Quality: An Overview." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1: 345–351. doi:10.5194/isprsannals-II-3-W5-345-2015.
- Barron, C., P. Neis, and A. Zipf. 2014. "A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis." *Transactions in GIS* 18 (6): 877–895. doi:10.1111/tgis.12073.
- Billen, R., and E. Clementini. 2005a. "Introducing a Reasoning System based on Ternary Projective Relations." In *Developments In Spatial Data Handling 11th International Symposium on Spatial Data Handling*, edited by P. Fisher, 381–394. Berlin, Heidelberg: Springer.
- Billen, R., and E. Clementini. 2005b. "Semantics of Collinearity among Regions." In *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops*, edited by R. Meersman, Z. Tari, and P. Herrero, Vol. 3762, 1066–1076. LNCS. Berlin, Heidelberg: Springer.
- Billen, R., and E. Clementini. 2006. "Projective Relations in a 3D Environment." In *Geographic Information Science: 4th International Conference, GIScience 2006*, edited by M. Raubal, H. J. Miller, A. U. Frank, and M. F. Goodchild, Vol. 4197, 18–32. LNCS. Berlin, Heidelberg: Springer.
- Bishr, M., and K. Janowicz. 2010. "Can We Trust Information? The Case of Volunteered Geographic Information." In *Towards Digital Earth Search Discover and Share Geospatial Data Workshop at Future Internet Symposium*, edited by A. Devaraju, A. Llavés, P. Maué and C. Kefler, Vol. CEUR-WS Vol. 640. ISSN 1613-0073. <http://ceur-ws.org>
- Bishr, M., and L. Mantelas. 2008. "A Trust and Reputation Model for Filtering and Classifying Knowledge about Urban Growth." *GeoJournal* 72 (3–4): 229–237. doi:10.1007/s10708-008-9182-4.
- Clementini, E. 2013. "Directional Relations and Frames of Reference." *GeoInformatica* 17 (2): 235–255. doi:10.1007/s10707-011-0147-2.
- Clementini, E., S. Skiadopoulos, R. Billen, and F. Tarquini. 2010. "A Reasoning System of Ternary Projective Relations." *IEEE Transactions on Knowledge and Data Engineering* 22 (2): 161–178. doi:10.1109/TKDE.2009.79.
- D'Antonio, F., P. Fogliaroni, and T. Kauppinen. 2014. "VGI Edit History Reveals Data Trustworthiness and User Reputation." In *Proceedings of the 17th AGILE Conference on Geographic Information Science, Connecting a Digital Europe through Location and Place*, edited by J. Huerta, S. Schade and C. Granell. Castellon, Spain: AGILE Digital Editions. 03–06 June, 2014. ISBN: 978-90-816960-4-3
- Dai, C., D. Lin, E. Bertino, and M. Kantarcioglu. 2008. "An Approach to Evaluate Data Trustworthiness Based on Data Provenance." In *Secure Data Management: 5th VLDB Workshop, SDM 2008*, edited by W. Jonker and M. Petković, Vol. 5159, 82–98. LNCS. Berlin, Heidelberg: Springer.
- Degrossi, L. C., J. P. De Albuquerque, R. D. Santos Rocha, and A. Zipf. 2018. "A Taxonomy of Quality Assessment Methods for Volunteered and Crowdsourced Geographic Information." *Transactions in GIS* 22 (2): 542–560. doi:10.1111/tgis.2018.22.issue-2.
- Egenhofer, M. J. 1989. "A Formal Definition of Binary Topological Relationships." In *Foundations of Data Organization and Algorithms: 3rd International Conference, FODO 1989*, edited by W. Litwin and H. J. Schek, Vol. 367, 457–472. LNCS. Berlin, Heidelberg: Springer.
- Egenhofer, M. J., and D. M. Mark. 1995. "Modelling Conceptual Neighbourhoods of Topological Line-Region Relations." *International Journal of Geographical Information Systems* 9 (5): 555–565. doi:10.1080/02693799508902056.
- Elwood, S., M. F. Goodchild, and D. Sui. 2013. "Prospects for VGI Research and the Emerging Fourth Paradigm." In *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*, edited by D. Sui, S. Elwood, and M. F. Goodchild, 361–375. Dordrecht: Springer.
- Fan, H., A. Zipf, Q. Fu, and P. Neis. 2014. "Quality Assessment for Building Footprints Data on OpenStreetMap." *International Journal of Geographical Information Science* 28 (4): 700–719. doi:10.1080/13658816.2013.867495.
- Fogliaroni, P., and E. Clementini. 2015. "Modeling Visibility in 3D Space: A Qualitative Frame of Reference." In *3D Geoinformation Science: The Selected Papers of the 3D GeoInfo 2014*, edited by M. Breunig, M. Al-Doori, E. Butwilowski, P. V. Kuper, J. Benner, and K. H. Haefele, 243–258. Cham: Springer.
- Fonte, C. C., V. Antoniou, L. Bastin, J. Estima, J. J. Arsanjani, J. C. L. Bayas, L. M. See, and R. Vateva. 2017. "Assessing VGI Data Quality." In *Mapping and the Citizen Sensor*, edited by G. Foody, L. See, S. Fritz, P. Mooney, A. M. Olteanu-Raimond, C. C. Fonte, and V. Antoniou, 137–163. London: Ubiquity Press.
- Freksa, C. 1991. "Conceptual Neighborhood and its Role in Temporal and Spatial Reasoning." In *Decision Support*

- Systems and Qualitative Reasoning*, edited by M. Singh and L. Travé-Massuyès, 181–187. Amsterdam: North Holland.
- Girres, J. F., and G. Touya. 2010. “Quality Assessment of the French OpenStreetMap Dataset.” *Transactions in GIS* 14 (4): 435–459. doi:10.1111/j.1467-9671.2010.01203.x.
- Goodchild, M. F. 2007. “Citizens as Sensors: The World of Volunteered Geography.” *GeoJournal* 69 (4): 211–221. doi:10.1007/s10708-007-9111-y.
- Goodchild, M. F., and L. Li. 2012. “Assuring the Quality of Volunteered Geographic Information.” *Spatial Statistics* 1: 110–120. doi:10.1016/j.spasta.2012.03.002.
- Haklay, M. 2010. “How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets.” *Environment and Planning B: Planning and Design* 37 (4): 682–703. doi:10.1068/b35097.
- Hartig, O. 2009. “Provenance Information in the Web of Data.” In *Proceedings of the 2nd Workshop on Linked Data on the Web (LDOW2009)*, edited by C. Bizer, T. Heath, T. Berners-Lee and K. Idehen, vol. CEUR-WS 538. ISSN 1613-0073. <http://ceur-ws.org>
- Helbich, M., C. Amelunxen, P. Neis, and A. Zipf. 2012. “Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata.” In *Geoinformatics Forum 2012: Geovisualization, Society and Learning*, edited by T. Jekel, A. Car, J. Strobl, and G. Griesebner, 24–33. Berlin, Offenbach am Main: Herbert Wichmann-Verlag
- Hernández, D., E. Clementini, and P. Di Felice. 1995. “Qualitative Distances.” In *Spatial Information Theory: A Theoretical Basis for GIS*, edited by A. Frank and W. Kuhn, Vol. 988, 45–57. LNCS. Berlin: Springer-Verlag.
- Keßler, C., and R. T. A. De Groot. 2013. “Trust as a Proxy Measure for the Quality of Volunteered Geographic Information in the Case of OpenStreetMap.” In *Geographic Information Science at the Heart of Europe*, edited by D. Vandembroucke, B. Bucher, and J. Cromptvoets, 21–37. Cham: Springer.
- Keßler, C., J. Trame, and T. Kauppinen. 2011. “Tracking Editing Processes in Volunteered Geographic Information: The Case of OpenStreetMap.” In *Proceedings of Workshop on Identifying Objects, Processes and Events in Spatio-Temporally Distributed Data (IOPE 2011), Workshop at COSIT'11*, edited by A. Galton, M. Worboys, M. Duckham and J. Emerson, unpublished. Belfast: Maine. 12–16 September 2011.
- Lodigiani, C., and M. Melchiori. 2016. “A PageRank-based Reputation Model for VGI Data.” *Procedia Computer Science* 98: 566–571. doi:10.1016/j.procs.2016.09.088.
- Mark, D. M. 1993. “Human Spatial Cognition.” In *Human Factors in Geographical Information Systems*, edited by D. Medyckyj-Scott and H. M. Hearnshaw, 51–60. London: Belhaven Press.
- Martella, R., C. Kray, and E. Clementini. 2015. “A Gamification Framework for Volunteered Geographic Information.” In *AGILE 2015 Geographic Information Science as an Enabler of Smarter Cities and Communities*, edited by F. Ba, C.ão, M. Y. Santos, and M. Painho, Vol. 217, 73–89. LNCS. Cham: Springer.
- Mezzetti, N. 2004. “A Socially Inspired Reputation Model.” In *Public Key Infrastructure. EuroPKI 2004*, edited by S. K. Katsikas, S. Gritzalis, and J. López, Vol. 3093, 191–204. LNCS. Berlin, Heidelberg: Springer.
- Miller, G. A. 1995. “WordNet: A Lexical Database for English.” *Communications of the ACM* 38 (11): 39–41. doi:10.1145/219717.219748.
- Mooney, P., and P. Corcoran. 2012. “Characteristics of Heavily Edited Objects in OpenStreetMap.” *Future Internet* 4 (1): 285–305. doi:10.3390/fi4010285.
- Mooney, P., P. Corcoran, and A. C. Winstanley. 2010. “Towards Quality Metrics for OpenStreetMap.” In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, edited by D. Agrawal, P. Zhang, A. El Abbadi and M. Mokbel, 514–517. New York: ACM Press.
- Moratz, R., and M. Ragni. 2008. “Qualitative Spatial Reasoning about Relative Point Position.” *Journal of Visual Languages & Computing* 19 (1): 75–98. doi:10.1016/j.jvlc.2006.11.001.
- Neis, P., D. Zielstra, and A. Zipf. 2012. “The Street Network Evolution of Crowd-sourced Maps: OpenStreetMap in Germany 2007–2011.” *Future Internet* 4 (1): 1–21. doi:10.3390/fi4010001.
- Randell, D. A., and A. G. Cohn. 1989. “Modelling Topological and Metrical Properties of Physical Processes.” In *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning (KR'89)*, edited by R. J. Brachman, H. J. Levesque and R. Reiter, 357–368. San Francisco, CA: Morgan Kaufmann.
- Raymond, E. S. 1999. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Sebastopol, CA: O'Reilly Media.
- Richter, K. F., and S. Winter. 2014. *Landmarks: GIScience for Intelligent Services*. Cham: Springer International Publishing.
- Senaratne, H., A. Mobasheri, A. L. Ali, C. Capineri, and M. Haklay. 2017. “A Review of Volunteered Geographic Information Quality Assessment Methods.” *International Journal of Geographical Information Science* 31 (1): 139–167. doi:10.1080/13658816.2016.1189556.
- Spielman, S. E. 2014. “Spatial Collective Intelligence? Credibility, Accuracy, and Volunteered Geographic Information.” *Cartography and Geographic Information Science* 41 (2): 115–124. doi:10.1080/15230406.2013.874200.
- Sztompka, P. 1999. *Trust: A Sociological Theory*. Cambridge: Cambridge University Press.
- Tarquini, F., and E. Clementini. 2008. “Spatial Relations between Classes as Integrity Constraints.” *Transactions in GIS* 12 (SUPPL. 1): 45–57. doi:10.1111/tgis.2008.12.issue-s1.
- Tarquini, F., G. De Felice, P. Fogliaroni, and E. Clementini. 2007. “A Qualitative Model for Visibility Relations.” In *KI 2007: Advances in Artificial Intelligence*, edited by J. Hertzberg, M. Beetz, and R. Englert, Vol. 4667, 510–513. LNCS. Berlin, Heidelberg: Springer.
- Wand, Y., and R. Y. Wang. 1996. “Anchoring Data Quality Dimensions in Ontological Foundations.” *Communications of the ACM* 39 (11): 86–95. doi:10.1145/240455.240479.
- Zielstra, D., and A. Zipf. 2010. “A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany.” In *Proceedings of 13th AGILE International Conference on Geographic Information Science*, edited by M. Painho, M. Y. Santos and H. Pundt, 1–15. Guimarães, Portugal: AGILE Digital Editions. ISBN: 978-989-20-1953-6