Trust as a Proxy Measure for the Quality of Volunteered Geographic Information in the Case of OpenStreetMap

Carsten Keßler and René Theodore Anton de Groot

Abstract High availability and diversity make Volunteered Geographic Information (VGI) an interesting source of information for an increasing number of use cases. Varying quality, however, is a concern often raised when it comes to using VGI in professional applications. Recent research directs towards the estimation of VGI quality through the notion of trust as a proxy measure. In this chapter, we investigate which indicators influence trust, focusing on inherent properties that do not require any comparison with a ground truth dataset. The indicators are tested on a sample dataset extracted from OpenStreetMap. High numbers of contributors, versions and confirmations are considered as positive indicators, while corrections and revisions are treated as indicators that have a negative influence on the development of feature trustworthiness. In order to evaluate the trust measure, its results have been compared to the results of a quality measure obtained from a field survey. The quality measure is based on thematic accuracy, topological consistency, and information completeness. To address information completeness as a criterion of data quality, the importance of individual tags for a given feature type was determined based on a method adopted from information retrieval. The results of the comparison between trust assessments and quality measure show significant support for the hypothesis that feature-level VGI data quality can be assessed using a trust model based on data provenance.

C. Keßler (✉) · R. T. A. de Groot
Institute for Geoinformatics, University of Münster, Münster, Germany
e-mail: carsten.kessler@uni-muenster.de

R. T. A. de Groot
e-mail: rta.de.groot@gmail.com

D. Vandenbroucke et al. (eds.), *Geographic Information Science at the Heart of Europe*, Lecture Notes in Geoinformation and Cartography,
DOI: 10.1007/978-3-319-00615-4_2,

1 Introduction

Professional geographic information is collected according to established standards, which allows the provider to guarantee levels of data quality, yet also results in high costs and sparse updates. Volunteered Geographic Information (VGI) has become an attractive alternative for use cases where professional geographic information is too expensive, not available for a theme,[1] or when timely updates are required (Goodchild 2007). In order to enable data consumers to get the best from both of these two data sources, a method is required that reliably filters VGI based on its quality. This approach would enable “cherry picking” on the data consumer’s side, enabling them to identify high-quality features and leave features aside that seem problematic. At the same time, such a method would support the mapping community in spotting features that might need improvement.

Previous research has mostly focused on the *overall* quality of VGI datasets, e.g., by analyzing the average positional accuracy based on a dataset with guaranteed data quality. Such an overall analysis of a dataset, however, does not allow for quality assessments of single features in the dataset. Moreover, it requires access to the ground truth data that the data consumer is trying to do without. To overcome these problems, a model to assess the quality of VGI based on its provenance has been proposed (Keßler et al. 2011a, b; Mooney and Corcoran 2012b), where each feature’s editing history is analyzed in order to assess its quality. These quality assessments are also referred to as *informational trust* (Bishr and Janowicz 2010) to denote the degree to which a data consumer can trust the information about the feature. Trust is a central principle in (real-world) social networks and has been shown to play a central role in online communities (Golbeck 2005). The trust analysis is based on intuitive notions such as the *many eyes principle* (Haklay et al. 2010), where the quality is more likely to be higher if more people have worked on a feature, as well as patterns of revisions such as vandalism or *edit wars* that indicate quality problems. In previous work, we have proposed a provenance vocabulary to annotate edits and enable the computation of a trust measure (Keßler et al. 2011a).

This chapter presents a practical application of this trust-based approach, along with an evaluation on a set of features selected from the OpenStreetMap (OSM) dataset for the city of Münster, Germany. The hypothesis for our work is that parameters derived from OSM provenance data determine a feature’s trustworthiness, which is an indicator for data quality and therefore, trustworthiness and data quality are correlated. We test this hypothesis by carrying out two parallel assessments: one based on trust indicators derived from the OpenStreetMap history, and the other one based on high-quality reference data collected in a field survey. The outcome of both datasets is then tested for statistical correlation.

The remainder of the chapter is organized as follows: The Sect. 2 provides an overview of related work. Section 3 describes the study area. Section 4 analyzes

---

[1] See http://wheelmap.org, for example.

the trustworthiness of the features in the test dataset, followed by an analysis of their data quality based on a field survey in Sect. 5. Section 6 provides a comparison and statistical analysis of the correlation between the results of the two methods, followed by concluding remarks in Sect. 7.

2 Related Work

At the time of writing of this chapter, the OpenStreetMap community consisted of more than a million mappers who had generated more than 1.7 billion nodes, more than 160 million ways, and more than 1.7 million relations between them.[2] With the growing number of applications building on OSM, data quality has become an issue that has already been addressed in a number of articles. Haklay (2010) provides one of the first comprehensive studies, comparing OpenStreetMap data in the UK with government data provided by Ordnance Survey. Focusing on motorway objects, he found that about 80 % of the features from both classes overlap. Neis et al. (2012) present a similar study for the German street network by comparison with commercial data provided by TomTom and find that the overlap is at about 91 % (as of June 2011). Zielstra and Zipf (2010) show that the OpenStreetMap dataset converges towards commercial datasets in terms of completeness. A statistical comparison of junction points in OSM and commercial data from TeleAtlas by Helbich et al. (2010) also supports the statement that OSM presents a suitable alternative for commercial data. Koukoletsos et al. (2012) introduce an automated feature matching method for comparative studies focusing on positional accuracy such as the ones mentioned above. While the results for positional accuracy in OpenStreetMap are largely positive, a comprehensive study by Mooney and Corcoran (2012c) shows that most of the quality problems seem to lie in the thematic data, i.e., in the tags assigned to the features.

The International Organization for Standardization (ISO) defines data quality of geographic information as *the difference between the dataset and a universe of discourse* (International Organization for Standardization 2002; Jakobsson and Giversen 2009), where the universe of discourse is the real world view defined by a product specification or user requirements. The ISO 19113 standard lists five data quality elements: completeness and consistency, as well as positional, temporal, and thematic accuracy. While these elements are generally provided as metadata for professional geographic information to allow an informed judgment of the dataset's quality, they are missing both for OpenStreetMap and VGI in general.

In the absence of quality metadata for VGI, trust has been proposed as a proxy measure for data quality (Bishr and Kuhn 2007). Trust can be defined as "a bet about the future contingent actions of others" (Sztompka 1999, p. 25) and is closely related to reputation, defined as the subjective perception of trustworthiness

---

[2] See http://www.openstreetmap.org/stats/data_stats.html

inferred from information about the historical behavior of somebody/something (Mezzetti 2004). The concept of trust commonly refers to interpersonal trust, but can be extended to informational trust through people-object transitivity (Bishr and Janowicz 2010). Bishr and Mantelas (2008) have successfully applied this principle in a scenario dealing with urban growth data. Closely related to trust is the concept of *credibility*, which is slightly broader, as it comprises someone's expertise (covering aspects of accuracy, authority and competence) in addition to trustworthiness (with aspects of reputation, reliability and trust; Flanagin and Metzger 2008). Artz and Gil (2007) point out that provenance is a key factor for trust on the web. The approach on provenance we take in this chapter is *data-oriented*, as the focus is on the origins of specific data items, instead of the processes that generate the data (Simmhan et al. 2005).

3 Test Dataset

The OpenStreetMap history dump[3] contains all nodes, ways, and relations that were ever added to OSM, along with each of its revisions. From this file, we extracted the features within the area of interest for our study, namely Münster's old town (*Altstadt*), shown in Fig. 1. The map is based on a shape file extracted from OSM for easier processing in GIS environments.[4] The boundaries for the area of interest are based on the city of Münster's district boundaries, which contains the Altstadt area as an official administrative district.

This area of the city was selected because of its high density of points of interest for sightseeing, shopping, dining, et cetera. It is thus made sure that the input dataset for our analysis contains enough features with a long enough editing history to allow for a meaningful analysis. At the same time, the area is small enough to allow for a field survey that collects data of the features on-site to create a ground truth dataset. However, re-mapping all features in the area of interest to create a ground-truth dataset is clearly not feasible. We therefore selected a subset of the features in the area of interest based on the number of revisions that the respective feature has undergone. The selection was made based on the number of revisions, as our trust assessment will be based on the provenance of the feature. This approach thus ensures that the test dataset has a rich enough input in terms of feature history. We decided that including up to 100 features in the field survey is feasible. This criterion is met if we include features that have at least 6 versions, which applies to the 74 features highlighted in Fig. 2.

[3] See http://planet.openstreetmap.org/planet/full-history/

[4] See http://download.geofabrik.de/osm/europe/germany/nordrhein-westfalen/muenster.shp.zip

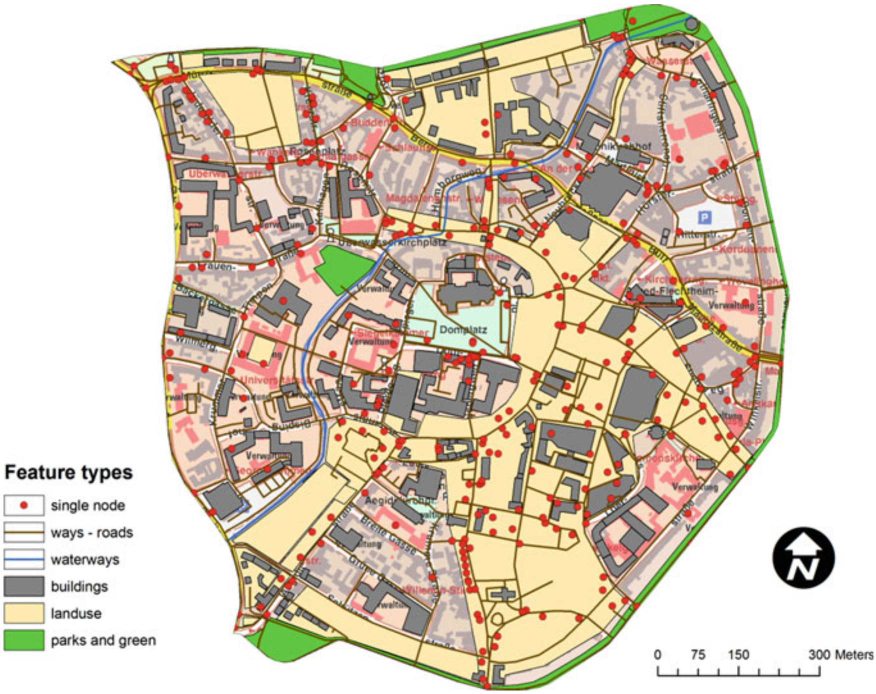

Fig. 1 Overview of Münster's old town based on the OSM data as of October 2011

4 Trust Assessment

This section introduces the trust assessment of the selected features. The parameters taken into account are introduced and the results of the assessment are discussed.

4.1 Parameters for Trustworthiness

First ideas on the parameters that influence trust assessments of features in OpenStreetMap have been discussed in Keßler et al. (2011a). However, not all proposed parameters can be applied to our test dataset, most notably user reputation. A measure for user reputation would require global knowledge of the OSM dataset to assess the user's experience (by counting the respective number of contributions) or even assessments of the quality of all features the user has been involved in. This is clearly not feasible for the small-scale test dataset used in this chapter, and is a complex research problem of its own. We are therefore taking the following parameters into account for our evaluation:

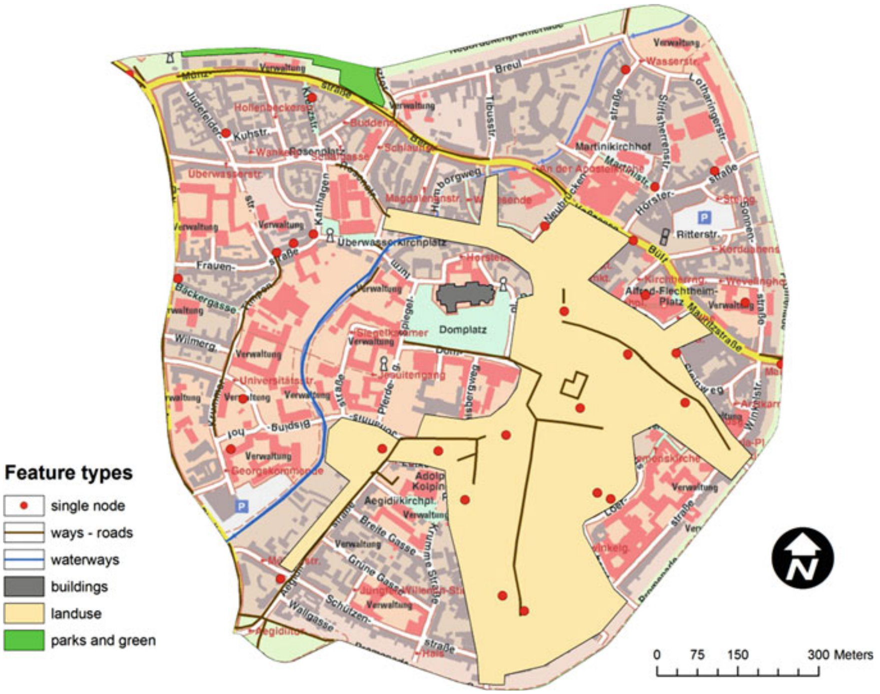

Fig. 2 Features with at least 6 versions highlighted

- **Versions**. As pointed out above, more versions essentially mean more provenance information to analyze. At the same time, they indicate that the feature has undergone a certain number of iterations with the goal to improve the feature's quality. We therefore assume that a higher version number adds to the feature's trustworthiness.
- **Users**. A higher number of users involved in creating a feature increases the trust measure, following the many eyes principle. Even though previous research found that the relation between number of users and data quality is not linear, we follow a linear approach here as the corresponding study by Haklay et al. (2010) has only taken positional accuracy into account and did not look at the quality of the thematic attributes.
- **Confirmations**. On top of the number of users that have been working directly on a feature, we also take *indirect* confirmations into account by looking at all revisions that have been made in the immediate vicinity of a feature *after* the last revision of a feature. The rationale is that it is very likely that someone who is editing features in a certain area also looks at features in the vicinity; therefore, we count all edits made within a 50 m buffer of a feature after its last revision as confirmations that increase trustworthiness.

- **Tag corrections**. Tags in OSM consist of a key and corresponding value.[5] Corrections occur when the value for a certain tag is changed, e.g., when the tag amenity = restaurant is changed to amenity = pub. We assume that this points to ambiguities in the feature classification and thus decreases trustworthiness.
- **Rollbacks**. A tag correction reverting a feature to its previous state is considered as a rollback and also decreases feature trustworthiness.

It is important to note that the parameters' influence on the trust measure as outlined above only covers the general case—there will always be exceptions to the rule. These exceptions, however, cannot be covered by generic statements, and can only be discovered by comparison with ground truth, which is not feasible in practice. We discuss in the following how well such a generic approach works.

4.2 Calculating Trust Assessments

For each of the parameters, we have created a classification into five equal intervals [1…5] that indicate a ranking from low to high trustworthiness. For those parameters with a positive influence on trustworthiness (numbers of versions, users, and confirmations, respectively), higher counts lead to a classification into a trustworthier interval. For the two parameters with a negative influence (corrections and rollbacks), *lower* counts lead to a classification into a trustworthier interval. As the combination of these parameters into a single trust assessment is still an open research question, we take a naïve approach here by assigning equal weights to all parameters. The overall results are then added up and classified into five equal intervals again for comparison with the results of the field survey.

Most features are almost equally distributed between classes two to four (65 out of 74), of which class four contains the highest number. Class one contains 8 features and class five only three. Overall, the trust measure suggests an estimation of the dataset as moderately trustworthy regarding its information quality, with only a low number of features reaching the highest level of trustworthiness. Figure 3 shows an overview of the trust assessments for the selected features.

5 Field Survey

In order to evaluate whether trust assessments bear results that realistically reflect feature quality, a reference dataset is required for comparison. This section describes how the reference dataset was collected in a field survey and how the data quality was rated.

[5] See http://wiki.openstreetmap.org/wiki/Tags for an overview of the tagging conventions the OpenStreetMap community has developed.

Fig. 3 Assessed trustworthiness of selected features based on numbers of versions, users, confirmations, corrections, and rollbacks

5.1 Thematic Accuracy and Topological Consistency

During the field survey, all selected features have been inspected on site, pictures were taken, and, if possible, people were asked to confirm the feature type in consideration to minimize bias. The following components were examined during this survey and are listed here in the order of higher to lesser importance regarding its influence on the quality of the thematic accuracy:

1. The correctness of the main tag: e.g. is this place a restaurant or a café?
2. The correctness of other tags that are described: for example, is the house number stated in OSM correct, or is the number of lanes of a street correct?
3. Is there any confusion or doubt about whether the description in OSM represents the feature in the right way:

- Unclearness about the nature of the feature; lack of description when it is not obvious what the function is (e.g., `information=guidepost:` is the guidepost about street information or historical information?)
- Doubt about the type within a main feature type (e.g., is it `high-way=primary` or `highway=secondary`?).

- The feature pointed to could be part of the whole feature instead of the feature itself (e.g., the entrance of parking lot could be marked as the parking lot feature)

Based on these criteria, the features were divided into four classes. Class 1 represents features of which the main tag does not correspond to what has been found in the field. Class 2 is assigned to features where other tags are incorrect. Class 3 contains features that have a shortcoming as described in the third point. And lastly, class 4 is assigned to those features of which the available information is fully correct. Out of the 74 features, 6 features ($\sim$8 %) were assigned class 1 (lowest thematic accuracy), 2 features (3 %) were assigned class 2, and 9 features ($\sim$12 %) were assigned class 3. For the remaining 57 features ($\sim$77 %), the available information is fully correct (class 4).

Positional accuracy in OpenStreetMap has already been extensively studied, as discussed in Sect. 2. For this reason, and for the lack of a straightforward method to create accurate positional information in our reference dataset, the selected features were tested for *topological* consistency. For all 74 features, it was checked whether they are qualitatively correctly positioned relative to the surrounding features. The results were positive throughout, except for one feature representing an information panel that was located on the wrong side of a street.

5.2 Information Completeness

The tagging guidelines documented in the OSM wiki provide recommendations which tags should be used for certain types of features, and describe how additional information such as opening hours can be added so that they can be automatically reused by applications building on the OSM data. However, due to the large number of potential feature types, the wiki does not provide lists of tags that should be provided for a certain type in order to provide complete information. We therefore had to come up with a method that determines importance of a specific tag for a given feature type. Based on the *term frequency—inverse distance frequency* measure (tf-idf) (Salton et al. 1975) that measures how characteristic a specific term is for a given document, we have defined an *inverse feature type frequency* (iff). It measures the general importance of a tag by dividing the total number of features in the dataset by the number of features with which this tag is associated. Taking the logarithm of this quotient generates values that indicate a higher relevance of the tag the closer it is to 0:

$$\mathrm{iff}(t) = \frac{\log|\mathrm{F}|}{|\{\mathrm{f}\colon \mathrm{t} \in \mathrm{f}\}|}$$

where $|\mathrm{F}|$ is he total number of features in the whole dataset and $|\{\mathrm{f}\colon \mathrm{t} \in \mathrm{f}\}|$ is the number of features in that set containing tag t. The *tag frequency—inverse feature type frequency* is then determined by:

$$\text{tf - iff}(t, f) = \text{tf}(t, f) * \text{iff}(t)$$

where tf is the tag frequency in the set of features that belong to the same feature type. The output of this calculation shows the relevance of a tag within a set of features of a certain feature type, also considering all the features in the whole dataset. Tags that are relatively unique to the feature type in consideration receive higher importance values than tags that are also commonly used to describe other feature types.

This approach turned out to be problematic for tags such as name that are relevant for a broad range of feature types. As they are not type-specific, their tf−iff measure is generally comparably low and would remove them from the list of important tags for a feature type. To solve this problem, we use the iff measure again, but now the denominators have a different meaning; $|F|$ is the total number of features within a feature type and $|\{f: t \in f\}|$ is the number of features with a certain tag within a feature type set. It turned out that this method filters out the most important and obvious tags per feature, even if they occur less frequently than in half of the features of a particular type. In a feature type like pub the tag smoking should be important. It occurred in less than half of all the pub features, but it is still incorporated in the set of 'obligatory tags' when determining its importance with the measure of general importance (iff). The values for both ways of importance determination were then normalized. We applied a threshold of 0.5 to identify the 'obligatory tags' for a given feature type.

In order to determine the information completeness for a specific feature, we checked each of the obligatory tags for presence in the feature and counted omitted tags. In order to stress the importance of tags that characterize a feature type, omissions of such tags (e.g., religion for churches) receive double weights.

5.3 Overall Quality

After preparing the individual quality element tests, linking them to the map IDs and classifying them into five classes (equal intervals), the outcomes of the three quality tests have been summed up and the summed values were reclassified again into five classes. 42 out of the 74 features (∼57 %) fully meet the quality requirements defined above. Generally, their theme is correctly described and their topological relations with respect to their surroundings are correct. The quality concerning information completeness is mixed; however, since each parameter of the quality assessment receives the same weight, the overall outcome is that many features are of high quality: 60 out of 74 features (∼81 %) are classified into the two highest quality classes. Figure 4 gives an overview of the observed data quality of the sample dataset.

6 Comparison and Results

This section compares the results of the trust assessment to those of the data quality observations collected in the field survey. We discuss the differences and provide a statistical correlation analysis.

6.1 General Observations

A visual comparison of the trust assessment shown in Fig. 3 and the observed data quality already shows that the features are generally deemed less trustworthy than indicated by the actual observed data quality shown in Fig. 4. A comparison including class difference is shown in Fig. 5. The visualization shows that the trust assessments generally underestimate the actual data quality: the mean quality of the features is at ∼4.2, whereas the mean trust value is at ∼2.8.

The trend across all features shown in Fig. 6, however, shows that the trust assessments generally follow the same pattern as the quality observations.

The salient high and low points of both measures revealed that low trust values were caused mostly by a low number of confirmations and the presence of tag

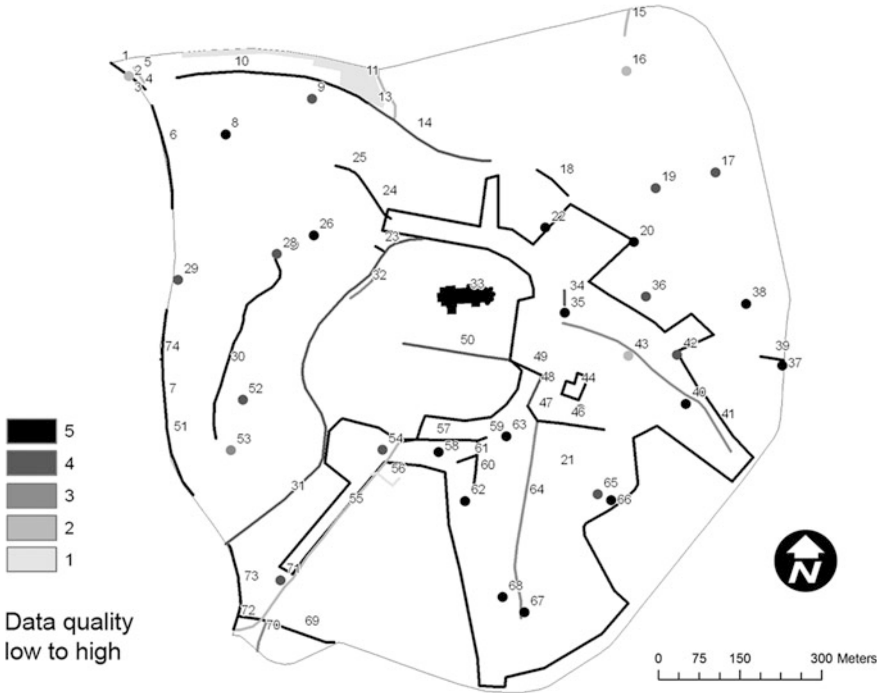

Fig. 4 Observed data quality of the sample dataset

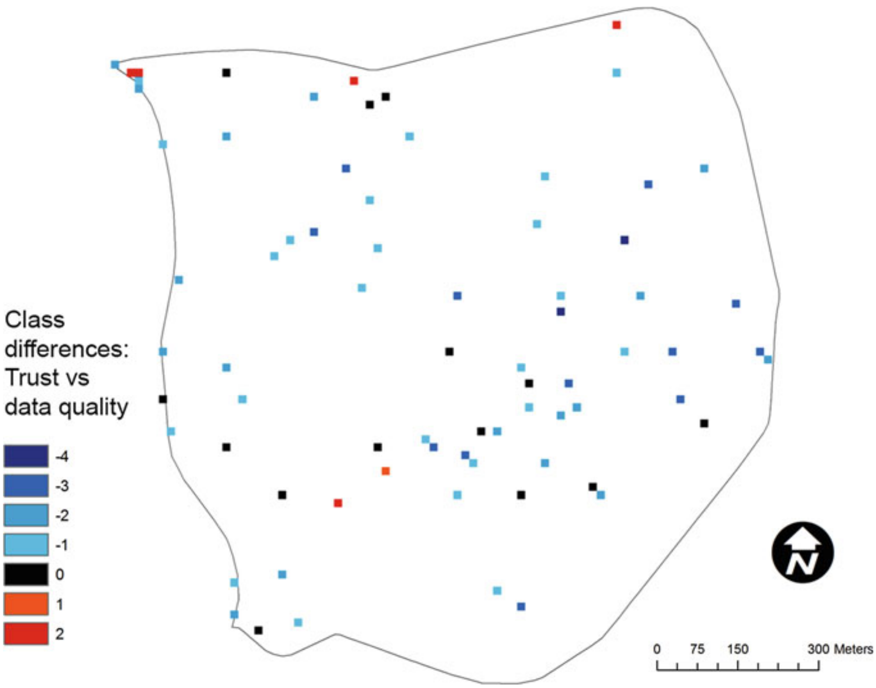

Fig. 5 Comparison of trustworthiness and data quality. *Blue colors* indicate that the trust assessments bear lower values than the observed data quality; vice versa for *red colors*

corrections or rollbacks, while the low quality values were often caused by a wrong feature type, in one case in combination with a high number of missing mandatory tags. This suggests that confirmations and rollbacks are the most distinctive indicators in our quality assessment.

Taking a closer look at the parameter values of the 20 features for which the classes of the two tests show no correspondence but rather opposition, reveals that a low number of versions, a low number of confirmations and the occurrence of tag corrections cause a low trust value for these features. In the rare case that the differences are in the opposite direction (high quality, but low trust), the theme of the feature was incorrect. There are only a few of these cases. Most of the big mismatches concern the first observation.

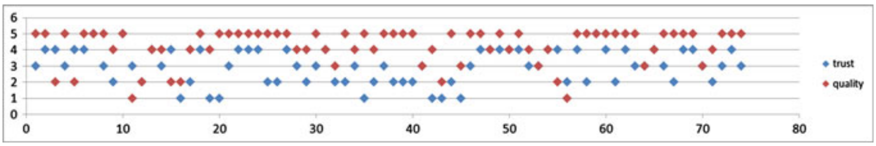

Fig. 6 Class value differences between the trust and quality measures; *the red dots* represent the quality measure results, *the blue* represent the trust measure results

The features with a high negative class difference generally have a higher number of tag corrections and slightly lower numbers of confirmations and users. Less emphasis on the tag corrections would increase the trust value of these features. Trying this out, however, resulted in more deviation from the trust value. This supports the model as is, but it also shows that one cannot always rely on counts of numbers of variables designated as trust parameters in order to get an indication of data quality.

6.2 Statistical Analysis

The general trend in Fig. 6 suggests a correlation between the trust assessment and the observed data quality. A statistical correlation analysis has been carried out to determine if there is an association between the two classifications, i.e., whether we can falsify our null hypothesis that there is no correlation between a trust assessment based on feature history, and the feature's observed data quality. Because the trust and quality measures both consist of a number of reclassifications and the quality measure includes a ranking, the whole set of trust and quality results does not have a clear numerical basis and is rather ordinal. We therefore applied Kendall's τ, a non-parametric rank correlation measure (Kendall 1938).

The total set of features was cleaned from outliers before the correlation measure was applied. The rationale here is that the 20 features not taken into account all show atypical behavior as discussed above. In this study, however, we are interested in whether using trust assessments as a proxy for data quality works at all, i.e., whether the trust assessments are able to predict the data quality in the common case. We thus leave a refinement of the trust assessment, including weights for the different parameters and the addition of new parameters such as user reputation, for future work.

Calculating Kendall's τ for the 54 features following the trend of the quality measure bears a positive correlation of $\tau \approx 0.52$, indicating a moderate correlation between the two measures. The p value for this correlation is close to zero (~ 0.00002), i.e., the correlation is highly significant. We can therefore reject the null hypothesis and accept the hypothesis that trust assessments and quality measures are correlated in a linear fashion. The statistical analysis thus suggests that provenance-based trust assessments do indeed work as proxies for data quality.

6.3 Discussion

Putting the result of the statistical analysis into context, several facts have to be taken into account. First, the model for trust assessment used in this chapter is very simple and does not take any of the finer details of trust models for VGI discussed in the existing literature into account. Specifically, the model does not take user reputation into account, which could not be computed in a meaningful way on the sample dataset. A realistic model of user reputation would have to take all edits a user has made into account and evaluate how the edited features have been treated by the community afterwards (e.g., by looking at revisions), potentially weighted by their respective local knowledge (van Exel et al. 2010). This would enable us to assess that user's reputation through an assessment of the quality of her edits. The required data for this, however, are not efficiently to retrieve through the OpenStreetMap API for a large number of users, so that we would have to bring the OSM history dump into a format that supports such queries first. As mentioned before, this is a research question of its own and we wanted to make sure that the idea of trust as proxy for data quality makes sense before we address user reputation in the next step.

Second, we have been using a straightforward combination of the five different parameters in this chapter by assigning all parameters the same weight. Similar to the user reputation, a detailed analysis is required in order to find out which of these parameters have a bigger influence on data quality. This could also be done in a study similar to the one presented in this chapter; however, a larger sample dataset would be required to be able to see how different weights affect the correlation of trustworthiness and data quality. The weighting of the features could then also address the fact that the trust measure in its current form generally underestimates the data quality.

Third, a more thorough study of the systematics behind the outliers that were not taken into account for the statistical analysis is required. A closer look at the outliers reveals that some of them were deemed too trustworthy with respect to their actual data quality because the trust assessment did not take information completeness into account, which could be included in future versions of the trust assessment in the same way we have included it in the quality measurement, i.e., by identifying the critical tags per feature type. This method obviously fails, however, if the feature type is not correctly assigned, which was also the case for some of the outliers.

Despite the small sample dataset used in our study, the statistical analysis shows that assessments of trustworthiness as a proxy for data quality are a topic that is worth pursuing in future research, which was our main motivation for the research presented in this chapter. In the following section, we will conclude the chapter and discuss directions for future work.

7 Conclusions

In this chapter, we have investigated the question whether trust assessments based on the provenance of features in OpenStreetMap can act as a proxy measure for data quality. To test this hypothesis, we have calculated trust assessments for 74 features in Münster's old town district based on a simple trust model. For comparison, a ground truth dataset has been collected in a field survey, for which we have defined a novel measure for tag importance in order to be able to measure information completeness. Based on a classification into equal intervals, the trust assessments and quality measurements per feature were tested for statistical correlation, which shows moderate, yet significant support for the hypothesis.

As discussed in the previous Sect. 6.3, this chapter is a first step towards meaningful quality assessments for Volunteered Geographic Information based on trust measures derived from feature provenance. This approach is novel in that it does not require ground truth data (except for the evaluation of the method itself) and can hence provide guidance for data consumers on an ad-hoc basis. Moreover, it bears the potential for new tools that support the community in spotting potentially problematic features that might need to be revised. In order to make trust as a proxy for VGI data quality operational, the weighting of the different parameters that influence trust needs to be analyzed, and user reputation needs to be taken into account. Moreover, the trust assessment needs to be extended to address the relatively large number of outliers that we have observed in our study.

The next steps in this research are therefore to collect a larger ground truth dataset to test different versions of the trust measure. The field surveys required to collect the ground truth data are very time-consuming and limited to a small area if, as in our case, performed by a single person. Crowd sourcing the quality assessment could hence be a useful approach to obtain a more diverse ground truth dataset, potentially through a smartphone app that lends ideas from gamification to motivate the users to participate. A larger ground truth dataset will allow us to address several open questions, including the actual importance and influence (negative or positive) of the parameters taken into account. This especially applies to the influence of tag corrections and rollbacks, which are currently treated as negative indicators. Moreover, additional and alternative approaches could be tested, such as using the number of change sets a feature is part of, instead of the number of versions. A systematic comparison of the test data against the best practices documented on the OpenStreetMap wiki could improve our understanding and rating of thematic accuracy and tag completeness. Putting these ideas into practice would also require an automation of the analysis, which was done largely manually for this chapter.

On the computational side, the user reputation needs to be addressed. If user reputation is handled as a function of a the quality of the respective user's edits, which in turn can only be assessed by the eventual development of the edited features, the computation becomes exponential and does not scale for large numbers of users. The user reputation assessment therefore needs to be driven by heuristics. User reputation is also an important aspect to address the quality assessments of features with a sparse editing history where analysis as discussed in this chapter is unlikely to bear meaningful results. In these cases, it is potentially more accurate to base the quality assessment solely on the reputation of the involved users, and eventually the social network relations between them (Mooney and Corcoran 2012a). User reputation is hence the most pressing issue to address in our future work.

References

Artz D, Gil Y (2007) A survey of trust in computer science and the semantic web. Web Semant 5:58–71

Bishr M, Janowicz K (2010) Can we trust information?—The case of volunteered geographic information. In Devaraju A, Llaves A, Maué P, Keßler C (eds) Towards digital earth: search, discover and share geospatial data 2010. Workshop at future internet symposium, Sep 2010

Bishr M, Kuhn W (2007) Geospatial information bottom-up: a matter of trust and semantics. In Fabrikant SI, Wachowicz M (eds) The European information society—leading the way with geo-information. Lecture Notes in Geoinformation and Cartography. Springer, Berlin Heidelberg, pp 365–387

Bishr M, Mantelas L (2008) A trust and reputation model for filtering and classifying knowledge about urban growth. GeoJournal 72(3–4):229–237

Flanagin AJ, Metzger MJ (2008) The credibility of volunteered geographic information. GeoJournal 72(3):137–148

Golbeck JA (2005) Computing and applying trust in web-based social networks. Ph.D. thesis, University of Maryland. Available from http://drum.lib.umd.edu/bitstream/1903/2384/1/umi-umd-2244.pdf

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

Haklay M (2010) How good is volunteered geographical information? a comparative study of OpenStreetMap and ordnance survey datasets. Environ Plann B, Plann Des 37(4):682–703

Haklay M, Basiouka S, Antoniou V, Ather A (2010) How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. Cartographic J 47(4):315–322

Helbich M, Amelunxen C, Neis P, Zipf A (2010) Investigations on locational accuracy of volunteered geographic information using OpenStreetMap data. GIScience 2010 Workshop on the role of volunteered geographic information in advancing science. Zurich, Switzerland. Available from http://ornl.org/sci/gist/workshops/2010/papers/Helbich.pdf

International Organization for Standardization (2002) ISO Standard 19113:2002: Geographic information—quality principles

Jakobsson A, Giversen J (2009) Guidelines for implementing the ISO 19100 geographic information quality standards in national mapping and cadastral agencies. EuroGraphics 2009

Kendall M (1938) A new measure of rank correlation. Biometrika 30(1–2):81–89

Keßler C, Trame J, Kauppinen T (2011a) Provenance and trust in volunteered geographic information: the case of OpenStreetMap. Poster presentation, conference on spatial information theory, 12–16 Sep 2011, Belfast, Maine, USA

Keßler C, Trame J, Kauppinen T (2011b) Tracking editing processes in volunteered geographic information: the case of OpenStreetMap. In Duckham M, Galton A, Worboys M (eds) Identifying objects, processes and events in spatio-temporally distributed data (IOPE), workshop at conference on spatial information theory 2011 (COSIT'11), 12 Sep 2011, Belfast, Maine, USA

Koukoletsos T, Haklay M, Ellul C (2012) Assessing data completeness of VGI through an automated matching procedure for linear data. Trans GIS 16(4):477–498

Mezzetti N (2004) A socially inspired reputation model. In: Proceedings of 1st European PKI Workshop. Lecture Notes in Computer Science, vol 3093. Springer, pp 191–204

Mooney P, Corcoran P (2012a) Who are the contributors to OpenStreetMap and what do they do?. In: Proceedings of 20th annual GIS research UK (GISRUK), Lancaster University, Apr 2012

Mooney P, Corcoran P (2012b) Characteristics of heavily edited objects in OpenStreetMap. Future Internet 4(1):285–305

Mooney P, Corcoran P (2012c) The annotation process in OpenStreetMap. Trans GIS 16(4):561–579

Neis P, Zielstra D, Zipf A (2012) The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. Future Internet 4(1):1–21

Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. Commun ACM 18(11):613–620

Simmhan YL, Plale B, Gannon D (2005) A survey of data provenance in e-science. ACM Sigmod Rec 34(3):31–36

Sztompka P (1999) Trust: a sociological theory. Cambridge University Press, Cambridge

van Exel M, Dias E, Fruijtier S (2010) The impact of crowdsourcing on spatial data quality indicators. GIScience 2010 workshop on the role of volunteered geographic information in advancing science. Zurich, Switzerland. Available from http://www.giscience2010.org/pdfs/paper_213.pdf

Zielstra D, Zipf A (2010) A comparative study of proprietary geodata and volunteered geographic information for Germany. AGILE 2010. In: The 13th AGILE international conference on geographic information science. Guimarães, Portugal