Reviewer A:

The authors compare official (bh-trans) and unofficial (waze) sources of data on urban events and traffic accidents. The article is well structured and understandable.

Details:

In general, the authors present good analyses and conclusion. However, in the Pag. 9, I think the follow conclusion is not the best one:

“This indicates a semantic discrepancy between regular citizens (Waze users) and police or transit officials, which has to be investigated further. “

Only 7% of BHTrans reports matched accidents reported on Waze. Thus, a possibility is that the accidents reported on Waze are minor accidents in fact. An interesting result would be to analyze only the matched accidents.

The suggested analysis was made. The paragraph to which the reviewer refers indicates that the conclusion refers to the matched accidents. We observed that, given an accident reported by both sources, what Waze users classify as "major" does not correspond to the more serious categories according to police or transit officials.

"**Among matched accidents,** Waze users only tend to classify as major severity accidents those officially classified as collisions involving pedestrians and rear-end collisions with victims (Table 2). In other cases, most accidents are deemed minor by Waze users. This indicates a semantic discrepancy between regular citizens (Waze users) and police or transit officials, which has to be investigated further."

Since we must assume that police and transit authorities have been trained to perform their functions, maybe the criteria they use does not correspond precisely to an individual Waze user's (passing by, maybe hurried) perspective of the same fact, thus the observation on the semantic discrepancy.

Reviewer B:

The paper addresses a relevant issue in the context of spatial database systems, geographic information systems and trajectory data. Besides, the contribution has a great possibility to be used to analyse real-world systems, that is, systems designed to manage and report traffic accidents.

The paper is well presented and has good contributions.

I strongly suggest that the authors include an additional paragraph in the end of the introduction to show clearly the improvements performed in this paper that represent the at least 30% of novelty compared with the paper of the Geoinfo 2016. I read the original paper of GeoInfo 2016 and I believe that this new version reaches the goal of improvement, but even so I suggest including the additional paragraph. For example, the new section 5.4 "Influence of Weather" should be pointed out as a novelty in this paper.
The paragraph was included at the end of introduction, as suggested. A note at the bottom of the first page indicates the paper's origin at GeoInfo and points out the expansions that were made.

Minor suggestions:
- in the introduction: "
As with other words with excrescent suffix -st, amidst is generally considered synonymous with simpler amid, and amid is preferred by style guides on both sides of the Atlantic.[1] ".
Corrected, thank you for observing that.

Reviewer C:

The work proposed in this paper consists in comparing car accident records provided by BHTrans and data collected from the mobile application Waze. The results derived from such comparison are important for urban planning or street building strategies. The article in its present form, however, needs to further clarify its focus and contribution in comparison with existing related approaches, methods, algorithms, methodologies and with current database technologies.
This paper does not propose a novel integration algorithm or method, and does not claim to have achieved better results in a comparative analysis. As indicated in the Introduction, our objectives (Introduction section) include using existing integration techniques (spatial and temporal proximity, attributes/schema matching) to improve on the analysis of traffic accidents in urban areas. We revised a sentence in the Introduction to make our objectives clearer.

The study is motivated by the usual difficulty in obtaining reliable and timely official data, thus leading to a question on whether or not such data can be replaced by volunteered information. Furthermore, we set forth to question the validity of existing datasets, given the characteristics of data collection procedures.

No novel approach in techniques or methodologies is proposed, and our results indicate that the datasets are so discrepant that the best integration techniques in the world could not solve what is basically bias driven by a data collection problem. Neither do authorities report the entire set of accidents, nor do people involved in accidents always report them to the authorities, and nor do volunteer passers-by report all of them. The integrated dataset allows, anyway, analyses related to accident hotspots, spatiotemporal distribution and a few others. It is clear that there is no guarantee that the integrated dataset covers the universe of traffic accidents during the given period of time.

Anyone using sophisticated Computer Science and database analysis techniques on biased data are certain to draw incorrect and biased conclusions. Our study shows that this can happen even when using official data sources.

There still exists a gap between the title and the content since no integration model or data integration schemas are provided. In fact, only two different datasets were used in the analysis performed -not indicating that in fact the databases have been integrated.
The title mentions the integration of data sources, not a model for integrating data sources. The integration was achieved, combining attributes from both sources and presenting results from the integrated dataset. We added Figure 1 and Table I to make the results from the integration of both sources more evident.

Unfortunately, I found a few issues with the paper that are too crucial to warrant publication:

1) The work motivation is not indicated clearly. Both data collected voluntarily and officially obtained data can help in solving many problems. This is already known. Section 1 indicates that official data can be enhanced or confirmed by data produced by volunteers – but this is also not new. What is the real motivation of the proposed work?
Official data are usually hard to obtain, given the limitations imposed by authorities as to the access to data on accidents. Official agencies are ill-prepared to provide timely and anonymized data on accidents, and therefore one can only hope to obtain data after a long delay. This is reported in the paper, at the beginning of the introduction.

We observe that, in Brazil, such behavior by authorities is a rule, rather than an exception. The use of crowdsourcing as a replacement for official data has then become a research theme for our group. See reference [Smarzaro et al, 2017] for instance.

In this case, our initial goal was to assess whether official data on accidents could be replaced or complemented by crowdsourced data, as indicated in the paper's introduction. Given Waze's intensive and growing use in Brazil, along with its features dedicated to report traffic incidents, Waze has become an obvious choice for crowdsourced data on traffic, as opposed to Twitter profiles (used in previous work by our group, see reference [Ribeiro Jr et al, 2012]), which have nearly disappeared in recent months --- but may still be valuable.

In order to verify whether official data and crowdsourced data sources represent the same phenomenon, we set out to perform a comparative analysis, surprisingly finding a wide discrepancy between them. The integration of accidents from both datasets to create a broader view of accidents was an obvious step then. This is reported in the paper.

Of course official data can be enhanced or confirmed by volunteers, even when taking into consideration the usual reliability and coverage limitations in volunteered sources, and we agree that this is not new, as research on VGI dates back to 2007. However, the level of discrepancy we found gives this work a new motivation, as it has clearly shown that someone investigating accidents based on official data is getting at best a partial view of the phenomenon, as indicated in our conclusions. Volunteered data have also failed in covering the whole picture.

2) The proposed methodology is described informally and textually. How can the adopted methods be generalized to be applied to other datasets? For the verification of possible matches, for example, an algorithm should be given.
Could a formalism be used to better describe the methods of Section 4?
We added Algorithm I to make the method more easily understandable.

3) Authors claim that the datasets used in the comparative analysis are complementary to each other. In fact, the number of matching records found was very small. If so, why the datasets need to be integrated? Can they be used separately only?
The fact that there is small number of matches between both datasets reinforces the need to have both datasets integrated to have a broader view and understanding of the accidents

across the city. Figure 2a and 2b shows how different is the spatial distribution of the accidents.

In this case study, official data dated back several months, back to a time when Waze was not so widely used as it is today. It is possible that future repetitions of the study would find a higher match rate. Anyway, the integration of the datasets is meant to produce a broader view on traffic accidents, a view that cannot be obtained by using each source separately.

Also notice that it is not possible to claim that the integrated dataset is complete. The conclusions clearly show the limitations of both datasets, so future analyses need to be careful in their assessment of the traffic accidents problem.

4) Some sections do not provide any scientific contribution for the computer science area. For example, see Section 5.3. Also, authors illustrate their ideas by applying a descriptive analysis of the collected data, but this is not sufficient for me.
Clearly our intention is not to advance Computer Science per se, but foster the use of Computer Science tools and techniques in the direction of solving actual problems of our society. It would be much simpler to provide detailed algorithms for the integration of toy datasets, with no social or real-world consequence, but we chose to deal with real data and their problems. This is clearly an applied Computer Science project, as are many of the GeoInfo contributions. In our opinion, Computer Science needs more initiatives like this, instead of closing up against interdisciplinary work. Our group regularly collaborates with transportation and traffic engineering scholars, although they are not co-authors of this paper.

5) What is the impact of your work on the data integration quality? Have you thought about new methods, methodologies, models, approaches for data source integration and analysis?
The integration was defined specifically to the datasets considered. The generalization of integration methods is beyond the scope of this work. As you said, It was not on the scope of this work to generalize data integration methods. There is no way to determine the quality of integrated data in the absence of ground truth, i.e., a complete dataset of every accident that took place during the time frame of the study. We showed that neither dataset can be used as ground truth for any study.

6) The article is full of not-validated claims – for example: It is indicated that Waze users cannot provide a correct severity classification of accidents.
The paper contains an observation of the discrepancies found in the classification of accidents that were matched between the two datasets. The classification of accidents by police or traffic authorities is supposed to be more detailed and professionally made, assuming there are standards for those professionals to use. No such accuracy can be expected from volunteer contributions, and that is the adequate interpretation of what has been said in the paper. Also notice that a Waze user may have only a few seconds in

contact with an accident site, and that the tool conditions his/her response to a few options. This is also indicated in the text.

All claims in the paper are based on the analysis of the data and on our interpretation of what these data imply. We attempted no comparison with other sources for validation, nor do we claim that the same results would be obtained in data from other cities or organizations.

7) Table III contains some words written in Portuguese.
The Portuguese words are street names.

Reviewer D:

Very well written paper about the integration of reports on traffic accidents.

I see a major problem on this paper is its purpose. The abstract says that the work explores the potential for integrating the two data sources. The case study shows that only 7% of the accidents are reported in the two datasets.
Well, the small overlap between the two sources is a clear indication that neither is able to represent the entire phenomenon. Integrating the sources in order to obtain a much larger set of accidents is a clear result of the work. We initially expected a much larger overlap, but the 7% indicate problems in the data collection processes related to traffic accidents.

On section 5 you realized a detailed and interesting analysis of the data of the case study. but, for what purpose? For real-time travel information, it doesn't work due to the time delay of the official data. As potential users of this work rest travel planning activities of the governments. For this planning the inclusion of non-official information sources could be of interest. This point of the objectives of the work must be explored better.

The focus of the paper seems to be more on solving traffic problems than on avoiding accidents.
A paragraph has been added in section 5 to clarify this point. The main potential impact of the analyses we conducted is the realization that any studies on the mitigation of traffic accidents (i.e., investigating common causes, correlation to other urban phenomena, identifying accident hot spots, and others) should look at official data with care, due to subnotification and bias in official accident reports. We never claimed that the results or analyses could be important for travel delay analysis, or travel planning.

Your section 2 is not about RELATED WORK but is a continuation of the introduction. Only the paragraph beginning with "Recent initiatives" is about related work.
The first two paragraphs from the related work present studies on traffic accidents. Given our objective, such studies were important in corroborating our view on official data and their deficiencies. From there on other works are cited, related to the type of study and analyses we present. We think our related work section is correctly connected to the paper's objectives, as stated in the introduction.