Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2018-200-AC2, 2019

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Climate risks, digital media, and big data: following communication trails to investigate urban communities' resilience" *by* Rosa Vicari et al.

Rosa Vicari et al.

rosa.vicari@enpc.fr

Received and published: 22 January 2019

We would like to thank Reviewer 1 for his time to review our manuscript and to provide constructive comments and suggestions. We will address all issues raised in the critique and we believe that our manuscript will be much stronger after addressing these comments. Hopefully the implemented changes will satisfy his requirements. Here we would like to list our preliminary responses to the items raised by the Reviewer:

Reviewer 1: The work presents an interesting comparison of text analysis from news and Twitter, to identify urban resilience networks during flood events. The presented work and results are very interesting, but the paper needs to be organised differently

and more technical details are necessary. Finally, a deeper analysis on why this work is useful needs to be presented.

Authors: We appreciate that the Reviewer expressed his interest for the research presented in this paper. We agree with him that the manuscript could benefit of a better organisation of its contents, additions with technical and explanations on why this research is useful.

1.

R1: Section 2.2 (Data) should, in my opinion, come before the Methods section (2.1), as the applied methods are specific to the collected data.

A: Following the Reviewer's comment, we will invert the order of the two sections.

2.

R1: The Data section can be divided in 3 subsections to present the three datasets.

A: Following the Reviewer's comment, we will divide Data section in three subsections

3.

R1: The Methods section should be considerably expanded. Most of the methodology is actually presented later on in the paper, and should be in this section instead.

A: We thank the Reviewer for this suggestion, we will enrich the Methods section with information from the following sections and the details requested in comment #4.

4.

R1: It would be very helpful to explain how Gargantext algorithm works, what it is based on. A lot is said about what Gargantext can do, but what did you do with it? Use active voices and present the logical order of passages.

A: We agree with the Reviewer that further details on Gargantext algorithms would be useful. The first list of terms is automatically extracted by Gargantext algorithms

on the basis of their occurrency, compared to other occurrences that characterise all Gargantext database corpora, as well as on the basis of the co-occurrencies that characterise the specific corpus. Then, the network representations are computed on the basis of the Louvain modularity. The Louvain method for community detection is used to maximise the network modularity. A network with high modularity has dense edges between the nodes within modules and sparse edges between nodes belonging to different modules. The modularity maximisation involves two stages: first the small clusters are detected, then the nodes that belong to the same cluster are aggregated and a new network is produced whose nodes are the clusters. These operations are repeated until a maximum of modularity is reached and a clusters hierarchy is built. All other publicly available information on how Gargantext works is online in Gargantext documentation: https://iscpif.fr/gargantext/. Furthermore, we will specify in the paper that the tasks carried by the authors consisted of: 1) selecting and downloading the two corpora of press articles (news covering the 2016 Seine river flood and the news covering the 2015 Alpes-Maritimes flood) from Europresse in html format; 2) uploading the file on Gargantext; 3) selecting, among the terms automatically extracted by Gargantext from the two corpora, the most relevant terms and merge synonyms, declensions of terms and equivalent forms; 4) for each corpus analysis, keeping only those terms that occur at least 5 times; 5) verifying the daily distribution of terms referring to resilience solutions (through Gargantext Analytics view) and record the daily occurrence values in an Excel file to generate a histogram (fig. 1); 6) launching on Gargantext the network representation based on conditional distance among the selected list of terms; 7) launching, through the network visualisation engine, the algorithm that allows the strongly related nodes to be positioned close to each other; 8) selecting the option to visualise the terms corresponding to each node; 9) zooming in the network to observe all the nodes (even those with a small degree); 10) capturing a photo of all the networks; 11) extracting the network in .gexf format in order to analyse it with Gephi software; 12) importing the .gexf in Gephi and convert it in two Excel tables with the node degrees and the edge weight; 13) generating through Excel the figures presented

in the Supplement.

5.

R1: Pg 3 Line 24: can you compare these statistics with the general population statistics?

A: The authors thank the Reviewer for highlighting that this additional data are necessary. We will include these data in order to comprehend the differences between characteristics of the French population and of the Twitter users in France.

6.

R1: How do you actually access the data from news and Twitter? Do you use an API? A scraper method? Which search criteria did you use? How many tweets did you download? You need all these details for reproducibility of results. A reader should be able to replicate all your steps.

A: We agree with the Reviewer that further details are needed on the method that we employed to access the news and tweet data. Concerning the tweets, we have first selected the tweets through Twitter Advanced Searches on the basis of the following criteria: all tweets published between 28/05/2016 and 2/7/2016 and that contain at least one of the following hashtags: #crue, #crueparis, #crueseine, #inondation, #inondations, #pluies, #Seine. We then used a scraping tool, Dataminer (an open-source chrome extension software) that allows the conversion of HTML data that appear in the browser window into clean Excel table format. We first obtained 10073 tweets, this amount was reduced to 4497 after deleting the tweets referring to flood located in a different region than Île-de-France Region, the tweets that included the term #crue but referred to 'Motley Crue', 'uncooked food' or 'cruelty' in French.

Concerning the news, we accessed them through Europresse.com, a press online database that allows to select articles on the basis of keywords (in the title or in all the article), authors' name, language, type of media (frequency of distribution, geo-

graphical area of distribution, language, country), media name, publication dates. The selection criteria were: French press articles published from 15/05/2016 to 15/10/2016, with a title including the terms (“crue” or “inond*”) and (“Seine” or “Ile-de- 30 France” or “Paris” or “Région Parisienne”).

7.

R1: Pg 4 Lines 28-30: these details should be in the data section.

A: Following the Reviewer’s comment, we will move these details to the data section.

8.

R1: Pg 4 Line 29: are these logical and/or? Is the and only between “inond*” and “Seine”? If so write is as an equation with correct parentheses.

A: We thank the Reviewer for suggesting the use of parentheses to facilitate the comprehension of the selection criteria. Parenthesis will be included as follows: (“crue” or “inond*”) and (“Seine” or “Ile-de- 30 France” or “Paris” or “Région Parisienne”)

9.

R1: Figure 1b: I would remove this panel. The case study is not presented in the analysis and generates confusion.

A: We agree with the Reviewer that it is better to remove Fig. 1b in order to improve the clarity of the paper.

10.

R1: Pg 6 Lines 1-2: details about the zooming capabilities are not relevant.

A: Following the Reviewer’s comment, we will remove these details.

11.

R1: Pg. 8 line 2: the colours are not relevant. Too much attention is given to the

cluster colours, although this has been assigned without meaning. Please remove the sentence here and the colour references in the list below. It is also a limitation for colour-blinded readers.

A: We agree with the Reviewer that references to the colours in the text are not necessary. We will remove them.

12.

R1: Pg 8 Line 11: I would personally specify Social Impact. Similarly at line 18, I would call it Economic Impact.

A: We thank the Reviewer for suggesting these cluster names. We will use the title "Economic Impact" instead of "Affected market report", but we prefer to replace "Impact record" with "Impact on population and infrastructure" since some of the key terms included in this cluster refer to infrastructure.

13.

R1: No comment is done on the keyword “resilience”, root concept in this paper. Is it find by the Gargantext networks? Is it common?

A: Following the Reviewer’s remark, we will include the following comment: the key term "resilience" ("résilience" in French) was automatically extracted by Gargantext from the first corpus (news covering the Seine river flood) but its occurrence in was below 5. We suppose that in 2016 the term resilience was not popular in the media debate yet.

14.

R1: Figures S2.1 and S3.1: can you put all the keywords by the histogram? Just one out of two appears.

A: We thank the Reviewer for pointing at this inaccuracy. We will correct the two figures so that all the keywords are visible.

15.

R1: Pg 10 Lines 3-6: remove references to colours as they are not meaningful.

A: Following the Reviewer's comment, we will remove the references to colors.

16.

R1: Figure 3: There is plenty of terms outside the defined clusters. Why the Impact Record cluster does not involve the keywords "passengers", "interrompu" and "victims" which seem relevant and close in the network? What about all the terms in the central/low part?

A: Following the Reviewer's comment, we will specify that in the second network, as a consequence of the smaller term occurrences, the number of nodes and edges is limited. Hence, not all the clusters are meaningful and can be identified with a macro-theme. Concerning the three key terms mentioned by the Reviewer, they don't belong to the Impact Record cluster because they are not violet. Indeed Gargantext highlights with the same colour all the terms belonging to the same cluster.

17.

R1: Pg 12 note 4: this should be included as a reference.

A: Following the Reviewer's comment, we will include this reference.

18.

R1: Pg 12 Lines 28-31: please explain why the "most liked users" and the "most retwitted users" are relevant in this analysis. What do they tell us about resilience?

A: These data reveal which Twitter users are the most influential and have the capacity to shape the social perception of risks and of urban resilience. We thank the Reviewer for highlighting that this point was not clear, we will include this comment in the text.

19.

R1: Pg 13 Lines 4-5: Probably people prefer to retweet from official users/news rather than individuals for a reliability reason. You prefer to share info from an official source, rather than a person.

A: We thank the Reviewer for this pertinent remark that will be inserted in the manuscript.

20.

R1: The Sections 3, 4, and 5 are already Results. I suggest you create a Section "Results" after "Methodology", with subsections for each of the case studies. Sections 3,4, and 5 also contain a lot of dscussions as well, which I would move to the "Results and Discussion" section, which should be renamed "Discussion" only. This would greatly improve the clarity of the manuscript.

A: We agree with the Reviewer that these changes will improve the clarity of the paper. We will reorganise the information presented in Sect. 3, 4, 5 in a new "Results" section and in the "Discussion" section, as it is suggested by the Reviewer.

21.

R1: Pg. 16 Line 27: the word "metric" would imply numerical values, but here you present mostly qualitative analysis. Do you have any plan to present additional quantitative analysis?

A: We agree with the Reviewer that the term "metric" is not the most adequate, we will replace it with indicator. Additional quantitative analysis is planned as part of our future research.

22.

R1: A big question is not answered: what is this study useful for? What can we learn from all this analysis? Why is it helpful? Is there anything that we can do differently in the future because of what we have learned?

A: Following the Reviewer's remark, we will include a reflection based on the following answer: the results obtained through this research are relevant to gain a better understanding of the public opinion. These results will be beneficial for any urban resilience project in the Paris region. They will contribute to creating a better connection with the urban community and optimise the project impact trough dialogue and cooperation with the stakeholders. Furthermore, the methodology can be easily applied to other urban areas or different climate related stresses and shocks.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-200, 2018.