First the authors would like to thank the Editor and the Reviewers for their time and comments that helped to improve the manuscript. We addressed all issues raised in the critique and we believe that our manuscript is now much stronger. Hopefully the changes implemented will meet their requirements! Here we list our answers to the items raised by Reviewer 1 and Reviewer 2.

Answer to Reviewer 1

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 needed:
—A better organisation of its contents. We address this issue in our answers to comments #1, #2, #3, #7, #20;
—Some additions with technical details. We address this issue in our answers to comments #4, #5, #6, #16;
—Explanations on why this research is useful. We address this issue in our answers to comments #18, #22.

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 inverted 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 divided the Data section in the following three subsections:

2.1 Press articles on the 2016 Seine River flood
2.2 Press articles on the 2015 flood in the Alpes-Maritimes Department
2.3 Tweets on the 2016 Seine River flood

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 enriched the 'Methods "section with information from the sections 3, 4 and 5 of the first version of the manuscript. The new version of the 'Methods' section includes the following subsections:

3.1 Method implemented to analyse the press articles
3.1.1 Dataset extraction
3.1.2 Aggregated analysis
3.1.3 Network representation
3.1.4 Visual observation of the network
3.1.5 Quantitative analysis of the nodes and the edges
3.2 Method implemented to analyse the tweets on the 2016 Seine River flood
3.2.1 Extraction of the dataset
3.2.2 Aggregated analysis
3.2.3 Users' profile and behaviour

We also added further details on the methodology following comment #4 and #6. Please, see our answers to these comments.

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 the manuscript would benefit of further details on how Gargantext works and on what did the authors. We described these aspects in Supplement 1 'Details on how Gargantext works and step-by-step implementation'.

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 in order to comprehend the differences between characteristics of the French population and of the Twitter users in France. We added these data in Footnote 2 (pag. 3).

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 press and tweet data.

Concerning the tweets, we included these details in Sect. 2.3: (from p. 4, line 20 to p. 5, line 2) and in 2.2.2.1 (p. 7, lines 21–24).
Concerning the press articles, we included more details on how Europresse works in Sect. 2.2.1.1 (p. 6, lines 6–8).

7.

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

A: Following the Reviewer's comment, we moved all the details concerning the selection criteria of the press articles and tweets to the 'Data' section (see Sect. 2.1, 2.2, 2.3).

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. Parentheses were included as suggested at p. 3, lines 16–17 and at p. 4, lines 2–3.

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 and we removed 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 removed 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 removed 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 replaced the title 'Affected market report' instead of 'Economic Impact', but we preferred to replace 'Impact record' with 'Impact on population and infrastructure' since some of the key terms included in this cluster refer to infrastructure. We made these changes in Fig. 2, Fig.3, p. 11 (lines 3 and 5), p. 13 (lines 9 and 10), p. 17 (line 4).

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 included comments on this point in Sect. 4.1 (p. 8, lines 29–31), Sect. 5.1 (p. 16, lines 21–23).

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 corrected the two figures (S3.1 and S4.1 in the supplements to the new manuscript) 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 removed these references to colours.

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, in Sect. 4.3.1 (p. 13, lines 3–4) we explain why some terms are outside the defined clusters.
Concerning the three key terms mentioned by the Reviewer, they don't belong to the 'Impact Record' cluster because they are not violet. In order to make sure that this point is clear, we specified in Sect. 3 (p. 6, lines 2–3) that '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 have included the note as the reference (Climaps, 2013) corresponding to 'Climaps: Reading the state of climate change from digital media, available at: http://disq.us/t/1gj2hci (last access: 18/03/2019), 2013.'

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: We thank the Reviewer for highlighting that this point was not clear, we included the following explanation in Sect. 3.2.3 (p. 8, lines 10–11): 'Indeed, 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'.

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. We inserted it in Sect. 5.3 (p. 17, lines 25–26).

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. As it is suggested by the Reviewer, we reorganised the information presented in Sect. 3, 4, 5 in a new 'Results' section and in the 'Discussion' section that include the following subsections:

4 Results
4.1 Aggregated analysis of the press articles
4.2 Graph representation based on the articles on the Seine River flood
4.2.1 Visual observation of the graph
4.2.2 Quantitative analysis of the nodes and the edges
4.3 Graph representation based on the articles on the Alpes-Maritimes flood

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 replaced it with 'indicator' in the 'Conclusions and perspectives' section. We also specified that additional quantitative analysis is planned as part of our future research at p. 19, lines 12–13.

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 included the following reflections in the 'Conclusions and perspectives' section (p. 19, lines 4–9).

Answers to Reviewer 2

Reviewer 2: This paper describes a study on the content analysis of the press and tweets about floods in France. The topic is relevant to the journal and the study is of interests to international readers. I support its publication subject to the following improvements (mainly clarifications):

23)

R2: The study has used Gargantext and Gephi systems. Please explain why they were chosen? Are there any alternative systems that could also be used?

A: We agree with the Reviewer that this choice should be explained and we made related additions in Sect. 3 (p. 5, lines 18-21), Sect. 3.1.5 (p.7, lines 16-18) and in Supplement 1.

24)

R2: In "when few documents are deleted from the corpus or few nodes are removed from the network." Do you mean "a few" instead of "few"?

A: Thank you for pointing at this grammar mistake, we corrected it (in Supplement 1 on the new manuscript version).

25)

R2: "excluding BTS"? what is "BTS"?

A: We agree with the Reviewer that this term should be explained. We have included this information in Footnote 1 (p.3).

26)

R2: In 'A corpus of 761 articles was first selected through Europresse archives', why did you choose Europresse? Any other alternative sources?

A: We agree with the Reviewer that this choice should be explained in the manuscript. We have included this information in Sect. 3.1.1 (p. 6, lines 8–11).

27)

R2: 'the following criteria: French press articles published from 15/05/2016 to 15/10/2016,' why did you choose the 5 month duration?

A: Following the Reviewer's comment, we added the following information in Sect. 2.1. (p. 3, lines 17–19).

28)

R2: In Figure 1, any explanations on the no values?

A: We agree with the Reviewer: we added the label 'n° of published terms per day' beside the y-axis to facilitate the comprehension of the figure.

29)

R2: In 'Table S1: Keywords related to flood resilient solutions', please explain the keywords in English and elaborate how they were selected (most readers cannot understand France, so it is useful to show the flood resilient solutions in English).

A: Following the Reviewer's comment, we included in the Supplements an English translation of the selected terms (Tab. S3) and we explain why this selection was made in Sect. 3.1.2 (p. 6, lines 17–18) and Sect. 3.1.3 (from p. 6, line 31 to p. 7, line 3).

30)

R2: In Figure S2.1, please show both the French and English terms.

A: Following the Reviewer's comment, we have translated all the French terms in the Supplements.

31)

R2: In 'The corpus of tweets covering the Seine River flood of June 2016 was extracted through "Twitter Advanced Search" (twitter.com/searchadvanced).' Any alternative sites? Why this?

A: We agree with the Reviewer that this choice should be explained. We included this information in Sect. 3.2.1 (p.7, lines 21–26).

32)

R2: 'The selection criteria were a time span (from 28/05/2016 to 2/7/2016)' why is such a duration chosen?

A: We made an addition to explain this choice in Sect. 2.3 (p. 4, lines 21–23).

33)

R2: There are many grammatical errors. Please check through the whole manuscript to remove them (e.g., 'it is possible to quickly navigating through'); 'This is probably due to the higher newsworthiness that events in the French capital have in comparison to those occurring in the rest of the country.' etc. . .).

A: Thank you for pointing at these mistakes, we checked the whole manuscript.

Climate risks, digital media, and big data: following communication trails to investigate urban communities' resilience

Rosa Vicari[1], Ioulia Tchiguirinskaia[1], Bruno Tisserand[2] and Daniel Schertzer[1]

[1]Hydrology Meteorology and Complexity Laboratory, École des Ponts ParisTech, Marne-la-Vallée, Champs-sur-Marne, 77455, France

[2]Veolia Research and Innovation, Chemin de la Digue, Maisons Laffitte, 78600, France

Correspondence: Rosa Vicari (rosa.vicari@enpc.fr)

Abstract. Nowadays, when extreme weather affects an urban area, huge amounts of digital data are spontaneously produced by the population on the Internet. These 'digital trails' can provide an insight on the interactions existing between climate-related risks and the social perception of these risks. According to this research 'big data' exploration techniques can be exploited to monitor these interactions and their effect on urban resilience. The experiments presented in this paper show that digital research can bring out the key issues in the digital media, identify the stakeholders that can influence the debate and, therefore, the community's attitudes towards an issue. Three corpora of Web communication data have been extracted: press articles covering the June 2016 Seine River flood; press articles covering the October 2015 Alpes-Maritimes flood; tweets on the 2016 Seine River flood. The analysis of these datasets involved an iteration between manual and automated extraction of hundreds of key terms; aggregated analysis of publication incidence and key term incidence; graph representations based on measures of semantic proximity (conditional distance) between key terms; automated visualisation of clusters through Louvain modularity; visual observation of the graph; quantitative analysis of its nodes and edges. Through this analysis we detected topics and actors that characterise each press dataset, as well as frequent co-occurrences and clusters of topics and actors. Profiling of social media users gave us insights on who could be the opinion makers"on Twitter. Through a comparison of the three datasets, it was also possible to observe how some patterns change over time, in different urban areas and in different digital media contexts.

1 Introduction

This paper presents a study on how digital media represent urban resilience during extreme weather and in the following weeks. The approach proposed here aims at exploiting the huge amount of Web data that are produced during and after a climate risk. Analysis of 'big data' corpora drawn from digital media has been already employed in research fields related to communication on climate risks. A series of studies analyse and compare the existing controversies on climate change in the web sphere (Niederer, 2013; Rogers and Marres, 2000; Chavalarias, 2015; climaps.eu, last access: 18 May 2018). Other research works investigate the use of social media during crisis due to climate hazards (Palen et al., 2010; Morss et al., 2017; Bruns et al. 2012; Lanfranchi et al., 2014; Gaitan et al., 2014; J.C. Chacon-Hurtado, 2017).

rosa 18/3/19 19:26
Deleted: …icari[1], Ioulia …chiguirin ... [1]

rosa 18/3/19 18:53
Deleted: –sur–

rosa 18/3/19 19:21
Deleted: "…igital trails'"…can provi ... [2]

rosa 18/3/19 11:51
Deleted: web …eb data that are prod ... [3]

The originality of our approach lies in the fact that it is framed in the context of research on urban resilience assessment. For this reason, it aims at contributing to the comprehension of the interactions that exist among different urban resilience drivers.

According to Mangalagiu, one of the challenges of the 21st century is giving the attention to the 'complex and causal linkages between human, technological, environmental and global biophysical systems' (Mangalagiu et al., 2012, p. 2). In our view, quantifiable variables facilitate the investigation of the relations between different physic-environmental and socio-economic components of urban systems. Furthermore, quantitative indicators are helpful to cross-compare different locations and time points. A huge variety of quantitative indicators are proposed in the literature on different resilience assessment approaches (Cutter et al. 2008; 2010; UN/ISDR, 2008; Resilience Alliance, 2010; Keating et al., 2014). In this paper, the focus is put on those quantitative data that can be used to investigate the social representation of climate risks. Digital media are a source of quantitative information that can be automatically extracted and analysed through computer-aided exploration techniques such as advanced text mining and graph representation. A quantitative analysis of digital communication patterns easily leads to an evaluation of how different space and time variables affect these patterns. It also facilitates the analysis of how communication trends and other resilience drivers (e.g. an environmental factor) mutually influence each other. When these correlations exist, they are a necessary basis to understand how social perception of climate risks affects urban resilience.

Besides examining these methodological challenges, we also intend to contribute to the comprehension of the social perception of urban resilience to climate risks through digital media. We present an analysis of three datasets in Sect. 3, 4 and 5: the press articles covering the June 2016 Seine River flood; the press articles covering the October 2015 Alpes-Maritimes flood; the tweets on the 2016 Seine River flood. We discuss an initial analysis of the topics and actors mentioned in the media texts, as well as of the thematic subsets and term co-occurrences that characterise each dataset. We also compare the three datasets and reflect on how the debate changes over time, in different urban areas and in different media contexts.

2 Data

Online press articles and social media posts are second-hand data. Indeed, as it is discussed by Venturini (Venturini et al. 2014), the researcher cannot directly control the production of these data and he should question himself about their production context and process. For instance, news items publication follows a set of journalistic values, the so-called 'newsworthiness' (Boyd, 1994), that determine if and how much a story is important for a media outlet and its audience. An example of news value is 'the greater the drama, the greater its prominence in conversation': this kind of news, that is expected to get their audience talking, is considered more worthy than others. These values are translated in a hierarchy of news that will guide news programming. In the case of social media, the problem of the digital divide (i.e. Internet access, skills and usage inequality) leads the researcher to consider the socio-demographic characteristics of social media users. For example, when analysing tweets, it should be taken into account that in France

rosa 18/3/19 18:54
Deleted: due

rosa 18/3/19 19:21
Deleted: "

rosa 18/3/19 19:21
Deleted: "

rosa 13/3/19 17:32
Deleted: metrics

rosa 18/3/19 10:23
Deleted: network

rosa 18/3/19 11:34
Deleted: news

rosa 18/3/19 11:34
Deleted: news

rosa 18/3/19 11:56
Deleted: most prominent

rosa 18/3/19 10:30
Deleted: connections

rosa 19/3/19 16:00
Deleted: **2 Benefits and constraints of digital media analysis based on advanced text mining techniques**

rosa 19/3/19 16:00
Deleted: **.1**

the population over 45 years is not well represented by a sample of Twitter users, while the population with a university degree is overrepresented. Indeed, in France in 2017, 16% of Twitter users were over 45 years old and 40% of the users had a university degree (excluding BTS[1]) (blogdumoderateur.com, last access: 18 May 2018)[2]. According to Venturini 'digital traces are not natural items but artefacts created in a specific environment and with specific objectives'. However, this doesn't reduce their value because their publication process can be a source of information on the social representation of reality, for instance the media representation of climate-related threats. Even though digital communications make possible a more direct observation of social phenomena, these data need to be contextualised and interpreted.

rosa 18/3/19 18:41
Moved (insertion) [15]

rosa 18/3/19 18:41
Deleted: Blog du moderateur: blogdumoderateur.com

rosa 19/3/19 15:08
Deleted: .

rosa 18/3/19 18:41
Deleted:

2.1 Press articles on the 2016 Seine River flood

We picked the Seine River flood that occurred in June 2016 as a case study because of its prominent media impact. According to a search of French press articles on Europresse (europresse.com, last access: 18 May 2018), on the 3rd of June 2016 the press coverage of this flood event reached a peak of 310 articles published in one day (corresponding to 29 437 terms, as illustrated by Fig. 1a). This is a remarkable figure considering that the same French media published 591 articles in one day on Trump's victory on the 8th of November 2016. Media visibility influences public opinion, hence stakeholders' attitudes towards risks and disasters, and related resilience policies or projects[3]. Therefore this flood event is worth exploring from the urban resilience perspective.

The corpus of articles on the Seine River flood consists in 761 documents selected on the basis of the following criteria: French press articles published from 15/05/2016 to 15/10/2016, with a title including the terms ('crue' or 'inond*') and ('Seine' or 'Île-de-France' or 'Paris' or 'Région Parisienne'). We chose the duration of five months in order to obtain a corpus of press articles that were large enough for graph representation and because the first analysis was carried out in November 2016, five months after the Seine River flood of June 2016.

2.2 Press articles on the 2015 flood in the Alpes-Maritimes Department

[1] BTS (Brevet de Technicien Supérieur) is a French diploma of higher education that is obtained after two or three years of studies in a highly specialised field.

[2] According to the French National Institute of Statistics and Economic Studies (insee.fr, last access: 20/01/2019), in 2017, in France, 20.9% of the population between 25 and 65 years old had a university degree (obtained after more than two years of university studies) and 41% of the population was over 45 years old.

[3] Media contribute to our perception of reality (including risks and disasters) through selection and omission of information. For example the UK government reassurance campaign contributed to spread the mad cow disease, which resulted in millions of animals being destroyed and the deaths of 226 people. Furthermore, humans respond more forcefully to emotional appeals than to facts like in the case of the Indian Tsunami earthquake (2004): images and stories from tourists and the extreme language used by the media led to a higher donors' response compared to other disasters with more victims.

On the 3rd and 4th of October 2015 extreme rainfall caused river floods in the Alpes-Maritimes Department. Cannes, Antibes, Vallauris, Biot and Mandelieu-la-Napoule were the most affected municipalities. The press coverage of this flood event was more limited (286 articles over five months) in comparison to the Seine River flood (761 articles over five months), even though the first flood took a huge toll on human life with 20 deaths.

The corpus of articles on the Alpes-Maritimes flood includes 286 documents. We used the following criteria for the selection: French press articles from 15/09/2015 to 15/02/2015, with a title including the terms ('crue' or 'inond*') and a reference to at least one of the locations affected by the flood[4]. We decided to consider the same duration (five months) in the analysis of the first and the second press article corpora in order to facilitate a cross comparison between the two case studies.

2.3 Tweets on the 2016 Seine River flood

As mentioned above, the press select and prioritise the news worthy information, hence it defines the prominent topics and their organisation in thematic clusters. In this way, editors and journalists obviously influence the public perception of an extreme weather event, even though a two-way relationship exists between the press and the audience. This bond has been progressively fading since access to information has hugely increased in terms of variety and quantity, as a consequence of different factors, among others the development of public relations by non-journalistic organisations and the pervasive role of the Web sphere (Bucchi, 2013; Trench, 2008). In this context, a corpus of texts published on a social media deserves to be analysed and compared to the press articles corpus in order to have insight of the public perception of a flood event beyond the borders of the journalistic arena. The third dataset analysed in this paper is a corpus of tweets covering to the Seine River flood of June 2016. The choice to focus on Twitter (twitter.com, last access: 18 May 2018) is due to the fact that public authorities and citizens are increasingly using Twitter during natural disasters as a two-way early warning and information channel. According to Bruns and Liang (2012, p. 1), Twitter is particularly suitable for crisis communication, indeed with its 'flat and flexible communicative structures' any visitor can access public tweets: users that are not yet followers of the account that disseminates the information on a crisis event or even visitors that are not registered on Twitter. Furthermore, hashtags can be used by any

[4] The titles of the articles selected for the second case study have a title referring to at least one of the following locations: ("Alpes-Maritimes" or "Cannes" or "Antibes" or "Vallauris" or "Biot" or "Mandelieu-la-Napoule" or "Bouches-du-Rhône" or "Var" or "Vaucluse" or "Drôme" or "Siagne" or "Brague" or "Fréjus" or "Reyran" or "Vallauris-Golfe-Juan" or "Cagnes-sur-Mer" or "Le Cannet Mougins" or "Nice" or "Roquefort-les-Pins" or "La Roquette-sur- Siagne" or "Théoule-sur-Mer" or "Valbonne" or "Villeneuve-Loubet" or "Les Arcs" or "Brignoles" or "Cabasse" or "Callas" or "Camps-la-Source" or "Flassans-sur-Issole" or "Côte d'Azur" or "sud-est" or "Flayosc" or "Forcalqueiret" or "Fréjus" or "Méounes-lès-Montrieux" or "La Motte" or "Néoules" or "Puget-sur-Argens" or "La Roquebrussanne" or "Saint-Antonin-du-Var" or "Saint-Raphaël" or "Le Thoronet" or "Trans-en-Provence").

visitor to search for tweets on a specific topic. Such communicative structure facilitates fast, large-scale collection of information.

The corpus of tweets was selected on the basis of two criteria. The first one is the time span: all tweets were published from 28/05/2016 to 2/7/2016. We decided to consider one month duration because during the flood and in the following month the incidence of tweets was relevant and the portion of tweets referring to other floods (outside the Seine River Basin) was reduced. The second criterion was the presence of relevant hashtags: each tweet contains at least one of these hashtags '#crue', '#crueparis', '#crueseine', '#inondation', '#inondations', '#pluies', '#Seine'. As it is suggested by Bruns and Liang (2012), thanks to hashtags, it is possible to focus on those tweets where terms related to a flood are marked as important information. Geolocation was not used as a criterion because the sample of tweets would have been very small: few users provide such detail with their tweet (Bruns et Liang, 2012). We first obtained 10 073 tweets, the corpus was then refined and reduced to 4497, after deleting duplicates and the tweets that included the term #crue but referred to 'Motley Crue', 'uncooked food' or 'cruelty' (in French). In order to facilitate a comparison with the two previous corpora, the tweets that mentioned locations outside the Paris region were deleted as well.

3 Methodology

Stakeholders' perception of a controversial issue is a community characteristic (and a social impact when change occurs) that can be analysed through surveys, meetings and interviews. Surveys provide information on population attitudes at aggregated level, while interviews and meetings (e.g. focus groups) provide insights about how and why particular attitudes are developed at individual level or at small-group level. However, big data exploration techniques make possible to get beyond the dichotomy between the aggregated structure and the individual component when studying social connections (Latour, 2012). The following examples of analysis of digital texts illustrate how computer aided exploration techniques can be employed to gain insight both on the intensity and the quality of Web communications. Indeed, thanks to an automated big data exploration tool such as Gargantext (Chavalarias and Delanoe, 2017), it is possible to quickly navigate through huge masses of digital information, follow the connections among cultural contents (e.g. articles, blog posts, tweets), popular topics, the names of public figures or organisations.

A multitude of methods have been developed in the field of the scientometrics (Leydesdorff and Milojevic, 2012) to reveal the internal structure of cultural domains, through an automated analysis of data extracted from digital texts (such as authors, sources, documents, citations, links, references, terms, etc.). One of these methods, employed by Gargantext, is 'co-word analysis', a technique to map the semantic structure in a database that exploits the statistics on the frequency of co-occurrences between pairs of terms. We chose to use this open-source software to facilitate the replicability of the study and because Gargantext is unique in terms of ergonomics (see Supplement S1 for further details on how it works and on the required manual interventions). Furthermore, thanks to the collaboration between the HM&Co laboratory and ISC-PIF we benefited of support from the developers of Gargantext.

rosa 7/3/19 18:11
Moved (insertion) [1]

rosa 19/3/19 16:01
Deleted: 2.2

rosa 18/3/19 18:57
Deleted:

rosa 17/3/19 13:19
Deleted: allow

rosa 17/3/19 13:19
Deleted: ting

rosa 18/3/19 10:05
Deleted: allow gaining

rosa 18/3/19 13:01
Deleted: web

rosa 18/3/19 18:46
Deleted: gargantext.org,

rosa 18/3/19 13:06
Deleted: ing

rosa 18/3/19 13:06
Deleted: and

rosa 18/3/19 13:08
Deleted: ing

rosa 18/3/19 13:07
Deleted: and

rosa 18/3/19 13:06
Deleted: .

rosa 18/3/19 13:45
Formatted: Not Highlight

rosa 18/3/19 13:14
Deleted:

With Gargantext we can extract, automatically as well as manually, a list of key terms from a corpus of texts. This list of terms is then used by Gargantext to compute a graph representation. A 'graph' or a 'network', 'is a simplified abstract representation of a system, such as text database, that is meant to facilitate a smart navigation in the analysed system. It is a mathematical structure that represents a collection of interconnected objects. The objects are called 'nodes' (or 'vertices') and the connections are called 'edges' (or 'links') (Newman, 2010). The analysis of the graph leads to the comprehension of its single objects (i.e. nodes), the interactions (i.e. edges) among those components and the pattern of interactions (i.e. graph). In the graphs computed with Gargantext the nodes represent the key terms and the edges represent a co-occurrence relation between two terms. In this research we computed graph representations on the basis of the semantic proximity measure, between pairs of key terms, that is called 'conditional distance'. Conditional distance between two terms is the probability that these terms co-occur in the same meaning unit (e.g. a press article). Gargantext graphs are weighted: a value (called the 'weight' of the edge) from zero to one is assigned to each edge of the graph and it indicates the probability that two terms co-occur. Furthermore, Gargantext can display the degree of each node, a measure that indicates the number of edges connected to each node.
The nodes are assembled in cohesive subsets (or clusters) through a clustering algorithm, more specifically through Louvain modularity. Each cluster corresponds to a group of key topics and key actors that frequently appear in the same article. Gargantext highlights with the same colour all the terms belonging to the same cluster.

3.1 Methodology implemented to analyse the press articles

3.1.1 Dataset extraction

The two corpora of articles were extracted through Europresse.com, a press online database where articles can be selected on the basis of keywords (in the title or in all the article), authors' name, language, the type of media (frequency of distribution, geographical area of distribution, language, country), media name, publication dates. Europresse has the advantage that it is possible to export the press articles in a format that is compatible with Gargantext. Furthermore, Europresse gives access to press sources in different languages (unlike Argus de la Presse, another important press online archive). This characteristic opens the path to future research applied to other countries than French-speaking countries.

3.1.2 Aggregated analysis

A first histogram was created to illustrate the intensity of press coverage, expressed as the number of published terms per day, and how it evolved over one month.

The following step of the analysis consisted in selecting terms that correspond to a range of 'flood resilience solutions'. 'Flood resilience solutions' is understood here in a large sense, as any kind of solution that is implemented to reduce flood damages. Since the objective of the analysis was to comprehend how Paris resilience is represented by the media, we selected all terms referring to solutions to cope with flood risk.

rosa 18/3/19 09:55
Deleted: allows

rosa 18/3/19 09:55
Deleted: ing

rosa 19/3/19 15:23
Formatted: Font:12 pt

rosa 19/3/19 15:23
Formatted: Justified

rosa 18/3/19 10:24
Deleted: network

rosa 18/3/19 19:21
Deleted: "

rosa 18/3/19 19:21
Deleted: " or

rosa 18/3/19 10:24
Deleted: "graph"

rosa 18/3/19 10:24
Deleted: Network representation can be used to visualise which key terms co-occur in the same meaning unit (e.g. a press article or a social media post) and with what frequency. The key terms are represented by the "nodes" of the network, and the co-occurrency relation among these nodes is represented by the network "edges". Gargantext allows computing network representations that remain stable even when few documents are deleted from the corpus or few nodes are removed from the network. The nodes are assembled in cohesive subsets through a clustering algorithm, more specifically through Louvain modularity5. A weight is assigned to each node and to each edge: in the first case it corresponds to the node centrality, i.e. the number of edges associated to that node; in the second case it corresponds to the frequency of a connection between two nodes.

rosa 19/3/19 15:23
Formatted: Default Paragraph Font, Font:12 pt

rosa 19/3/19 15:23
Formatted: Font:12 pt

rosa 7/3/19 18:10
Deleted:

rosa 18/3/19 13:44
Formatted: Font:Italic

rosa 7/3/19 18:11
Moved up [1]: 2.2

rosa 7/3/19 18:10
Deleted: Data ... [4]

For each corpus, a first list was automatically established by Gargantext algorithms. These lists were then manually refined on the basis of the thematic relevance of each term (as a solution aimed to cope with a flood event) and finally merged. This resulting list (Tab. S2.1, translated in Tab. S2.2) was used to analyse the occurrence of the key terms in each corpus, except for the terms with fewer than five occurrences that were not considered in order to highlight the most frequent topics.

We used the key terms list to compare the total number of published terms per day with the number of published key terms per day referring to the topic of flood resilience solutions. The comparison between the overall corpus with the sub-group of key terms aimed to monitor how the quality of the contents evolved over time. Further insights can be obtained through a comparison between the two histograms based on different corpora of press articles and with the histogram based on a corpus of tweets.

rosa 12/3/19 20:48

Deleted: Figure 1b shows a similar histogram based on press articles covering the 2018 Seine River flood (21/01/2018 – 20/02/2018). In this case the portion of key terms referring to flood resilience solutions is reduced, since it varies between 0% and 5%. A

3.1.3 Graph representation

After analysing the key terms incidence in an aggregated manner, the second step of the analysis was aimed at representing a complex communication field and revealing the pattern of interactions that exist between different topics and public figures or organisations. The graph was based on an adjusted version of the key term list referring to flood resilience solutions. Indeed, for a better comprehension of the context, where these solutions were implemented, we added to the initial list the terms referring to Paris infrastructure and properties. Since another objective of this analysis was to identify the opinion makers, we also selected terms referring to stakeholders. Furthermore, synonyms ('inhabitants' and 'residents'), declensions of terms (e.g. 'scientist' and 'scientists') and equivalent forms (e.g. 'Établissement Public Seine Grands Lacs', the public institution managing the Seine River Basin, and 'EPTB Seine Great Lacs') were merged. Once the key term list was refined, a graph was generated on the basis of the conditional distance between pairs of key terms.

3.1.4 Visual observation of the graph

As it is suggested by Venturini (Venturini et al., 2014), we based the first analysis of the results on a visual observation of the graph and in particular of its clusters: which nodes belong to the same cluster and what the corresponding topics and actors are. According to Venturini's approach, each cluster can be associated with a macro-theme, chosen by the researcher, corresponding to an expression that sums up the terms in the cluster. In our research the macro-themes correspond to resilience management areas.

rosa 12/3/19 20:50

Deleted: colour of the cluster is automatically defined by Gargantext algorithm, while the

3.1.5 Quantitative analysis of the nodes and the edges

Further information on the graph can be obtained through a quantitative analysis: the values corresponding to the node degrees and to the edge weights can be easily extracted through Gephi, a graph visualisation software (gephi.com, last access: 18 May 2018). Gephi has the advantage of being an open-source software and of opening GEXF files, the format of the graph files produced by Gargantext. Furthermore, with Gephi it is possible to easily convert GEXF files in two Excel tables with the node degrees and the

edge weights.

3.2 Methodology implemented to analyse the tweets on the Seine River flood

3.2.1 Extraction of the dataset

The corpus of tweets covering the Seine River flood of June 2016 was selected through 'Twitter Advanced Search' (twitter.com/search-advanced, last access: 8 July 2018), a free and ergonomic service provided by Twitter. We then used the scraping tool Dataminer (an open-source Chrome extension software) to convert HTML data, that appear in the browser window, into clean Excel table format. By combining these two tools, it is possible to extract Twitter data, even after their publication and without any rate limit. An example of a different method to extract tweets is Twitter Search API, but in this case it is only possible to gather popular tweets published in the last seven days and there is a rate limit.

3.2.2 Aggregated analysis

rosa 13/3/19 13:15
Deleted:

An aggregated analysis of the sample of tweets was made through Gargantext, like for the article corpora. The term list (first row of key terms in Tab. S5.1, translated in Tab. S5.2) was established on the basis of the previous key term list and relevant new terms that are specific to Twitter jargon and also include English terms. The term resilience occurs fewer than five times also in this corpus. The occurrence of key terms was so limited that it was not possible to represent a significant graph, even after extracting terms referring to stakeholders and affected infrastructure. This is due to the limited number of characters that are allowed for each tweet (up to 140 in 2016) and that make the information essential, unless the tweet includes a link to an external webpage (the contents of which can't be automatically analysed through Gargantext).

Even if the thematic patterns of tweets could not be represented through a graph, we were able to push forward the aggregated analysis. We identified thematic groups of key terms and their frequency, as it was proposed by Vieweg et al. (2010).

3.2.3 Users' profile and behaviour

Besides this thematic analysis, the same sample of tweets can be used to investigate the behaviour of its users, their profiles and their interactions. We followed an approach which draws on the maps presented in the Climaps platform (Climaps, 2013). We identified the most liked, retweeted and active users in the sample. Indeed, these data reveal who are the most influential Twitter users that have the capacity to shape the social perception of risks and urban resilience.

rosa 18/3/19 18:31
Formatted: Not Highlight

We considered as the 'most active users' those accounts that published more than 10 tweets in one month. We identified as the 'most liked users' those accounts that received more than 50 likes per tweet in one month. We named those accounts that received more than 50 retweets per tweet in one month the 'most retweeted users'.

We then examined the area of activity of these users. This was established on the basis of

the description included in each user's account.

We then gave a closer look at the first five most active users. We aimed to observe if frequent tweeting had an impact in terms of popularity, i.e. in terms of likes and retweets.

4 Results

4.1 Aggregated analysis of the press articles

Due to a very high inflow of publications in few days, especially in the case of the Seine River flood, the information in Fig. 1 is presented in a semi-log plot to make the information clearer.

During Seine River flood the press coverage peak was reached on the 3rd of June 2016 (310 articles, corresponding to 29 437 terms) (Fig. 1a). In the case of the Alpes-Maritimes flood the peak of published terms per day (see Fig. 1d) is reduced (108 articles on the 4th of October, corresponding to 14 772 terms). As shown in Fig. 1d, after the maximum peak (on the 4th of October) the number of terms per day decreases less progressively than in the Paris region case study (Fig.1a). A peculiarity in Fig. 1d is that there are two small peaks on the 12th of October and then on the 28th of October.

316 key terms referring to flood resilience solutions (see Tab. S2.1 and S2.2) were extracted from the two corpora of articles: the corpus of articles covering the Seine River flood and the corpus of articles covering the Côte d'Azur flood. The key term 'resilience' ('résilience' in French) was automatically extracted by Gargantext but its occurrence was below five in both corpora, hence it was excluded from the list. The portion of listed key terms varies between 0% and 10% in the first corpus (Fig. 1a). In the second corpus the subset varies from 0% to 9% (Fig. 1d), which is close to the percentages of the previous case study.

Figure 1. Comparison between the total number of terms per day (light blue) and the number of terms per day referring to flood resilience solutions (dark blue) in a semi-log plot (based on three different datasets).

4.2 Graph representation based on the articles on the Seine River flood

4.2.1 Visual observation of the graph

The first graph (Fig. 2) includes 254 nodes, 445 edges and six clusters that can be associated with six macro-themes:

Figure 2. Graph representation computed with the press articles on the 2016 Seine River flood: co-occurrence graph computed on the basis of the measure of conditional proximity between pairs of key terms, in the corpus of articles covering the Seine River flood.

1. The cluster named 'Monitoring system' that brings together topics such as

rosa 19/3/19 16:06
Deleted: 3

rosa 19/3/19 16:06
Deleted: 3….1 Aggregated analysis ... [5]

rosa 11/3/19 16:55
Moved down [3]: Due to a very high inflow of publications in few days, the information in Fig. 1a) is presented in a semi-log plot to make the information more clear.
The press coverage peak was reached on 3rd of June 2016.
The portion of flood resilience solutions key terms varies between 0% and 10% in the first corpus (Fig. 1a).

rosa 13/3/19 16:04
Deleted:

rosa 11/3/19 16:55
Moved (insertion) [3]

rosa 11/3/19 16:56
Deleted: a)…is presented in a semi-lo ... [6]

rosa 11/3/19 17:22
Deleted: T

rosa 11/3/19 17:37
Deleted: 2016 …eine River flood and ... [7]

rosa 11/3/19 16:54
Moved (insertion) [2]

rosa 18/3/19 16:49
Deleted: first…revious case study. ... [8]

rosa 11/3/19 19:57
Moved down [4]: Figure 1. Comp ... [9]

rosa 11/3/19 19:57
Moved (insertion) [4]

rosa 18/3/19 19:26
Deleted: …. Comparison between ... [10]

rosa 19/3/19 16:06
Deleted: 3

rosa 11/3/19 19:54
Deleted: network

rosa 18/3/19 16:52
Formatted: Font:Bold

rosa 11/3/19 19:58
Deleted: Figure 2. Complex netwo ... [11]

rosa 11/3/19 19:58
Moved down [5]: <#>The clus ... [12]

rosa 13/3/19 16:04
Deleted: <#> ... [13]

rosa 11/3/19 19:58
Moved (insertion) [5]

rosa 12/3/19 20:51
Deleted: green macro-theme …l ... [14]

'sensors', 'data', 'modelling', 'observation', 'estimates', 'possible underestimations', as well as actors that include national authorities (the French Ministry of the Environment, the Minister of the Environment, Vigicrues) and researchers (IRSTEA, Vazken Andreassian);

2. The cluster named 'Warning system' that includes key terms such as 'red vigilance', 'warning', 'information', 'municipal plan', 'shelter', 'traffic' and actors, including government representatives (the French Minister of the Interior, the Prime Minister), a law enforcement institution (Prefecture), a transport company (SNCF), the national meteorological service (Météo France), an electric utility company (ERDF), vulnerable population (elders);
3. The cluster named 'Impact on population and infrastructure' that gathers topics such as 'rescue', 'closure', 'evacuation' of the 'museum', 'station', 'hospitals', 'transports', 'boats', 'cars', 'electric network', 'zouave' and actors such as local authorities (the Mayor of Paris, Departmental Council), cultural institutions ('Louvre', 'Grand Palais', National French Library), rescue services ('police', 'volunteers', 'civil protection', 'René Rabiant'), affected population ('tourists', 'tenants');
4. The cluster named 'Short-term response and insurance system' that brings together 'state of natural emergency', 'fire brigade', 'rescue services', 'stored art works', 'François Hollande', 'the National Assembly', 'inhabitants';
5. The cluster named 'Economic impact' with topics like 'damages', 'repair', 'farms', 'agricultural holdings', 'companies', 'market gardening', 'insurance' 'diagnosis' and actors that include the victims ('farmers', 'producers' 'victims', 'ship owners'), organisations representing them ('farmers' union') and insurers ('Bernard Spitz');
6. The cluster named 'Long-term solutions' includes topics like 'awareness raising', 'prevention', 'soil sealing reduction', 'La Bassée pilot project', 'memory', 'preparation', 'first pilot area', 'new structure', 'public debate', 'retention basin', 'reinforce', 'simulation'; and actors such as local authorities ('Regional Department of Environment and Energy', 'Métropole du Grand Paris', 'associations', 'public territorial agencies', the mayors of Saint-Maur-des-Fossés and Rueil-Malmaison, the Hydrology Director at the Public Agency of the Seine Grands Lacs Basin).

4.2.2 Quantitative analysis of the nodes and the edges

As it is shown in Figure S3.1, the nodes with the highest degree (degree> 30) concern warning and emergency management, especially management of public infrastructure, indeed they are located in the two biggest clusters (the 'Warning system' cluster and the 'Impact on population and infrastructure' cluster). For instance, the node with the highest degree is 'RER' (the Paris region commuter rail service) with 55 edges. Concerning the actors, law enforcement and rescue services appear as the actors in the debate with the highest degree.

In Figure S3.2 the most probable co-occurrences (i.e. the highest edge weights) are identified with the label of the corresponding couple of nodes. The figure highlights which terms tend to be paired; terms that concern the same risk management field (e.g.

rosa 18/3/19 19:21
Deleted: "…ensors'"… '"…ata' ... [15]

rosa 12/3/19 20:51
Deleted: light blue macro theme ... [16]

rosa 12/3/19 20:52
Deleted: turquoise macro-theme ... [17]

rosa 19/3/19 16:18
Formatted: Font:(Default) Times

rosa 19/3/19 16:18
Deleted: Impact report…"…that ... [18]

rosa 12/3/19 20:52
Deleted: dark blue macro-theme ... [19]

rosa 12/3/19 20:54
Deleted: yellow macro-theme … ... [20]

rosa 12/3/19 20:53
Deleted: light green macro-them ... [21]

rosa 19/3/19 16:06
Deleted: 3.…2.24…Quantitative a ... [22]

rosa 11/3/19 20:10
Moved down [6]: As it is shown in Figure S2.1, the most central nodes (degree > 30) concern warning and emergency management, especially management of public infrastructures, indeed they are localised in the two biggest clusters (the light blue cluster and the turquoise cluster). For instance, the most central node is "RER" (the Paris region commuter rail service) with 55 connections. Concerning the actors, law enforcement and rescue services appear as the most central actors in the debate.
In Figure S2.2 the most frequent co-occurrences (i.e. the highest edges weight) are identified with the label of the corresponding couple of nodes. The figure highlights that frequently coupled terms are: terms that concern the same risk management field (e.g. rescue services); terms that concern the same area of affected activities (agriculture or transports); terms that concern infrastructure located in the same flood prone area; two terms that refer to an action and the object to which this action is directed (e.g. « clubs » and « closing ») or a subject and an action performed by it (e.g. « trains » and « normally circulating »).

rosa 13/3/19 16:13
Deleted: .

rosa 11/3/19 20:10
Moved (insertion) [6]

rosa 18/3/19 19:26
Deleted: …2.…1, the most central ... [23]

rosa 18/3/19 19:26
Deleted: …2.…2 the most ... [24]

rescue services); terms that concern the same area of affected activities (agriculture or transports); terms that concern infrastructure located in the same flood prone area; two terms that refer to an action and the object to which this action is directed (e.g. 'clubs' and 'closing') or a subject and an action performed by it (e.g. 'trains' and 'normally circulating').

4.3 Graph representation based on the articles on the Alpes-Maritimes flood

4.3.1 Visual observation of the graph

Figure 3. Graph representation computed with the press articles on the 2015 flood in the Alpes-Maritimes region: co-occurrence graph computed on the basis of the measure of conditional proximity between pairs of key terms in the corpus of articles.

The second graph (Fig. 3) includes 104 nodes and 676 edges. As a consequence of the reduced key 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. However we identified the following four clusters with corresponding macro-themes:

1. A cluster of terms referring to the macro-theme 'Emergency management';

2. A cluster of terms related to the macro-theme 'Monitoring system and prevention';

3. A cluster of terms related to the macro-theme 'Reconstruction';

4. A cluster of terms related the macro-theme 'Impact on population and infrastructure'.

Except for the macro-theme 'Impact on population and infrastructure', the other themes are not equivalent to those identified in the first graph.

4.3.2 Quantitative analysis of the nodes and the edges

The nodes with the highest degree (Fig. S4.1) correspond to the terms 'deaths', 'missing' and 'victims'. Furthermore, the following actors are among the high degree nodes: the government (including the Prime Minister and the French President), the inhabitants, the rescue services (the volunteers, the police, the fire brigade), the Cannes mayor, celebrities and insurers.

The values corresponding to the edge weights (Fig. S4.2) show that the most probable co-occurrences have some similarities with the trends described in the previous case study: some couple of terms concern the same area of flood resilience management (such as forecasts, impact report, awareness raising and prevention or compensations for the victims); other pairs of terms can be identified as an action and a related object (e.g. an event cancellation). Two actors that are frequently coupled are CNRS (the French National Centre for Scientific Research) and Météo France (the French national weather service). Two other frequent edges connect scientific organisations (IRSTEA and University of Avignon) with the topic 'awareness raising'.

rosa 18/3/19 19:21
Deleted: « …lubs' »…and '« …losi ... [25]

rosa 19/3/19 16:06
Deleted: 3….5 … Press coverage o ... [26]

rosa 11/3/19 16:54
Moved up [2]: The peak of publis ... [27]

rosa 11/3/19 20:13
Deleted: The peak of published ter ... [28]

rosa 11/3/19 20:17
Moved (insertion) [8] ... [29]

rosa 18/3/19 19:26
Deleted: …. Network representati ... [30]

rosa 18/3/19 17:00
Formatted ... [31]

rosa 11/3/19 20:13
Deleted: with 104 nodes and 676 ed ... [32]

rosa 11/3/19 20:14
Moved down [7]: 1. A cluster of t ... [33]

rosa 13/3/19 16:54
Deleted: . ... [34]

rosa 11/3/19 20:14
Moved (insertion) [7] ... [35]

rosa 12/3/19 20:55
Deleted: turquoise subset …luster o ... [36]

rosa 12/3/19 20:55
Deleted: green subset …luster of ter ... [37]

rosa 12/3/19 20:55
Deleted: light blue subset …luster o ... [38]

rosa 12/3/19 20:56
Deleted: violet subset …luster of ter ... [39]

rosa 18/3/19 17:02
Formatted ... [40]

rosa 13/3/19 15:59
Deleted: Impact record…" ... [41]

rosa 18/3/19 19:21
Deleted: «

rosa 18/3/19 17:02
Formatted ... [42]

rosa 13/3/19 15:59
Deleted: Impact record …»… the ot ... [43]

rosa 11/3/19 20:14
Deleted: 4….3.2 Quantitative anal ... [44]

rosa 11/3/19 20:21
Moved down [9]: The nodes centr ... [45]

rosa 13/3/19 17:26
Deleted: . ... [46]

rosa 11/3/19 20:21
Moved (insertion) [9] ... [47]

rosa 15/3/19 13:05
Deleted: nodes centrality …Fig. S3. ... [48]

rosa 18/3/19 17:03
Deleted: s…weight … (Fig. S3….2) ... [49]

4.4 Aggregated analysis of the tweets on the Seine River flood

Figure 1c is based on a corpus of 7984 tweets and shows the number of terms published per day. The figure highlights that a term peak occurred on the 3rd of June 2016, like in the case of the press articles covering the same event. The dark blue columns in Fig. 1c represent the portion of terms referring to flood resilience solutions that varies between 0% and 3,5% of the total number of terms published per day.

Figure 4. Twitter coverage of the 2016 Seine River flood: key terms incidence aggregated in ten thematic categories. The key terms were extracted from the corpus of tweets.

As shown in Fig. 4, a major portion of key terms (3143 occurrences) consists in purely factual information, with references to the time and location of the flood event. The category 'flood resilience solutions' (1420 occurrences) includes a relevant portion of terms, as well as the 'stakeholders' category (1060 occurrences). However these are less frequent than terms describing the weather event (1506 occurrences) and its impact (1644 occurrences). On the contrary, the debate on the causes (72 occurrences) and risks (320 occurrences) is of little account.

4.5 Users' profile and behaviour in the Twitter coverage of the Seine River flood

Figure 5 is based on the following sample: 59 'most active users' (who published more than 10 tweets in one month), 43 'most liked users' (who received more than 50 likes per tweet in one month) and 58 'most retweeted users' (who received more than 50 retweets per tweet in one month).

Figure 5a highlights that individual Twitter accounts (i.e. accounts that aren't owned by an organisation but by a person) represent a relevant portion of the most active users as well as of the most popular users. Among the most active users, 37% own an individual account. The most liked users are characterised by a majority of individual accounts (65%). The percentage is reduced in the case of the most retweeted users: only 46% of them own an individual account. Hence, with a percentage difference of 19%, tweets from individuals generated less retweets than likes.

Figure 5b shows that the majority of users deal with weather forecasts (20%) or public administration/policy-making (19%), the next biggest area of activity gathers those users that are active in the field of journalism (14%). Information tweeted by public authorities and policy makers seems more frequent than information tweeted by rescue services (5%).

Figure 5. The users' behaviour (Twitter coverage of the 2016 Seine River flood): (a) percentage of individual profiles; (b) the area of activity of the most active users (59 users that published more than 10 tweets in one month); (c) the area of activity of the most liked users (43 users that received more than 50 likes per tweet in one month); (d) the area of activity of the most retweeted users (58 users that received more than 50 retweets per tweet in one month). These data were extracted from the

rosa 11/3/19 20:17
Moved up [8]: Figure 3. Network representation of the press articles on the 2015 flood in the Alpes-Maritimes Region.

rosa 11/3/19 20:19
Deleted: 5 Twitter coverage of the Seine River flood in 2016

rosa 18/3/19 17:04
Formatted: Font:8 pt

rosa 11/3/19 20:19
Deleted: 5.2 Aggregated analysis

rosa 11/3/19 20:25
Moved (insertion) [10]

rosa 18/3/19 19:26
Deleted: …. Twitter coverage of t ... [50]

rosa 13/3/19 17:27
Deleted: 5.3

rosa 11/3/19 20:29
Deleted: We identified the most active, liked and retweeted users in the sample. We consider as the "most active users" those accounts that…igure 5 is based on the ... [51]

rosa 11/3/19 20:25
Moved up [10]: Figure 4. Twitter coverage of the 2016 Seine River flood: key terms incidence aggregated in ten thematic categories.

rosa 18/3/19 17:09
Deleted:

rosa 11/3/19 20:30
Moved down [11]: Figure 5a highlights that individual Twitter accounts (i.e. accounts that aren't owned by an organisation but by a person) represent a relevant portion of the most active users as well as of the most popular users. Among the most active users, 37% own an individual account. The most liked users are characterised by a majority of individual accounts (65%). The percentage is reduced in the case of the most retweeted users: only 46% of them own an individual account. Hence, with a percentage difference of 19%, tweets from individuals generated less retweets than likes.

rosa 11/3/19 20:30
Moved (insertion) [11]

rosa 18/3/19 19:26
Deleted:

rosa 11/3/19 20:31
Deleted: Figure 5b presents the majority of users deal with weather forecasts (20% ... [52]

rosa 11/3/19 20:33
Deleted: Figure S5 : By highlighting the mean values and the upper bound[7] val ... [53]

tweet corpus and completed with information available on Twitter users' profile pages.

As shown in Fig. 5 c, the most liked users are journalists (21%), followed by users that operate in the field of public administration/policy-making (16%), art/historical monuments (16%) and transports (14%).

By looking at the areas of activity of the most retweeted accounts in Fig. 5d, we can notice similar trends to those presented in Fig. 5 c: the most relevant segment is the press (29%), followed by public authorities/policy-making (14%) and art/historical monuments (12%). The portion of tweets published by users that deal with transports is smaller (10%) than in Fig. 5 c with a percentage difference of 4%.

Lastly, in Fig. S6, by putting in evidence the mean values and the upper bound[8] values, we can observe that the retweets and likes of the five most active users follow similar patterns.

6 Discussion

6.1 Comparison of the three histograms

A comparison of the three histograms in Fig. 1 reveals that, in the press as well as in social media, the peak of publications per day always occurs on the day of the highest river discharge. This clearly suggests a correlation between a meteorological event and its social representation. A possible explanation could be that the press tends to rather focus on the immediate consequences of natural disasters (Houston et al., 2012). Twitter seems to follow the same trend as the press. Figure 1 also shows a noteworthy difference between the Alpes-Maritimes flood and the Seine River flood. The Seine River flood is marked by a slower decrease of press coverage, after the maximum coverage peak is reached. This kind of evolution could be a consequence of the very slow decrease of water levels of the Seine River, and it is probably reinforced by the high media visibility of Paris. Indeed, the events that occur in the French capital have a higher newsworthiness than those occurring in the rest of the country. Furthermore, the economic risks related to a flood event in Paris region are extremely high (OECD, 2014) since it is a densely populated area that represents a third of the national economy, and where companies' headquarters, national and international institutions are located; lastly, it is an important transportation node and one of the first tourist destinations in the world.

A peculiarity in Fig. 1d is that there are two small peaks: on the 12th of October this is caused by the press coverage of two rescued persons and two victims; on the 28th of October this is caused by the press coverage of the mayor of Cannes calling for help to the movie celebrities.

[8] After finding the interquartile range (IQR) and the upper quartile (Q3), the upper bound is calculated with the following formula : Upper bound = Q3+1.5*IQR.

rosa 11/3/19 20:33
Moved down [12]: As shown in Fig. 5c the most liked users are journalists (21%), followed by users that operate in the field of public administration / policy making (16%), art / historical monuments (16%) and transports (14%).

rosa 18/3/19 17:10
Deleted:

rosa 18/3/19 17:10
Deleted:

rosa 11/3/19 20:33
Moved (insertion) [12]

rosa 18/3/19 19:26
Deleted: / …olicy- …aking (16%), ... [54]

rosa 18/3/19 19:26
Deleted: / …olicy- …aking (14%) ... [55]

rosa 18/3/19 17:13
Formatted: Font:Not Bold, Not Highlight

rosa 12/3/19 15:43
Moved down [13]: The press coverage peak was reached on 3rd of June 2016; on the same day the Seine River discharge was highest.

rosa 11/3/19 20:36
Deleted:
3.2 Network representation ... [56]

rosa 13/3/19 17:37
Formatted ... [57]

rosa 12/3/19 15:54
Deleted: 3.2 Network representati ... [58]

rosa 13/3/19 17:37
Formatted ... [59]

rosa 12/3/19 15:43
Moved down [14]: Since the total ... [60]

rosa 12/3/19 12:06
Deleted: Results and d

rosa 13/3/19 17:29
Deleted: 6….1 Comparison of the ... [61]

rosa 12/3/19 15:43
Moved (insertion) [13]

rosa 12/3/19 15:51
Deleted: The press coverage peak w ... [62]

rosa 12/3/19 15:43
Moved (insertion) [14]

rosa 13/3/19 16:15
Deleted: four …hree histograms in F ... [63]

rosa 18/3/19 12:16
Formatted ... [64]

It should also be noticed that in Fig. 1c the maximum peak is not significant as in the histograms based on the press corpora. We suppose that Twitter can be influenced by the press agenda by it doesn't totally adhere to it.

The extraction of terms related to flood resilience solutions, through an iteration between manual analysis and automated text mining of the press corpora, revealed that the term 'resilience' occurs fewer than five times in each corpus: we suppose that in 2015 and 2016 the term resilience was not popular in the French media debate yet.

Through an aggregated analysis of the list of key terms, we identified thematic subsets in each dataset and we observed the distribution of the number of published key terms per day (Fig. 1). The comparison of the three datasets calls attention to the minor discussion on flood resilience solutions on Twitter if we compare it to the debate in the press. Indeed Twitter is conducive to a fragmented communication and it is typically used as an early warning system: to disseminate factual information on the time and location of a flood.

6.2 Comparison of the graph representations

A comparison between the first and the second case study reveals some interesting similarities and differences between the two. The second corpus is characterised by a smaller occurrence of its key terms that generates a graph with a reduced number of edges and nodes. The macro-theme 'Impact on population and infrastructure' appears in both graphs and it is probably a recurring topic in the press coverage of natural disasters. The other macro-themes don't occur in both graphs. This is indicative of a certain variability between two cases of floods in terms of the resilience levers that are covered by the press.

In the first corpus of articles, the debate on various levers for flood resilience is well developed and detailed. Furthermore, specific stakeholders can be associated with specific resilience levers. The nodes with the highest degree in the first graph (for instance 'RER' with 55 edges) let us suppose that the revolved around immediate flood impacts the Seine River flood, i.e. how it affected a relevant portion of the population in their daily life. An unexpected result is that some stakeholders involved in flood risk management in Paris are not visible in the media debate in 2016. It is also surprising that the press doesn't refer to nature-based solutions as an alternative to traditional defence solutions to cope with flood risk.

The observation of the second graph let us suppose that the important number of victims drew the media attention to the tragic consequences of the flood event and to emergency management. Indeed, we observed that the stakeholders corresponding to high degree nodes are the affected population and those organisations that were involved in rescue activities, economic compensation and commemoration of the victims. Nevertheless, the most frequent edges reveal that part of the press debate revolved around the resilience solutions proposed by the scientists and the national weather service.

6.3 Analysis of the tweets

rosa 18/3/19 17:35
Deleted: A list of …erms related to ... [65]

rosa 18/3/19 10:02
Deleted: An …hrough an aggregate ... [66]

rosa 12/3/19 16:14
Deleted:
The next stage of the analysis involved involved network representation applied to the 2016 Seine River flood case study and the 2015 Alpes-Maritimes case study. Thanks to an observation of the thematic clusters it was possible to gain a qualitative insight on how the debate on flood resilience is structured. A quantitative evaluation of the most central nodes and of the most frequent connections in the network enabled us to push forward the analysis: we identified the most prominent topics and actors and which of these are often coupled together. ... [67]

rosa 18/3/19 17:43
Deleted: The next stage of the analysis involved network representation applied to the 2016 Seine River flood case study and the 2015 Alpes-Maritimes case study. Thanks to an observation of the thematic clusters it was possible to gain a qualitative insight on how the debate on flood resilience is structured. A quantitative evaluation of the most central nodes and of the most frequent connections in the network enabled us to push forward the analysis: we identified the most prominent topics and actors and which of these are often coupled together. … comparison betw ... [68]

rosa 18/3/19 19:08
Deleted: to

rosa 18/3/19 17:46
Formatted: Not Highlight

rosa 12/3/19 17:40
Deleted: impact …onsequences of t ... [69]

rosa 13/3/19 17:30
Deleted: 6.…3 Tweets a ... [70]

The aggregated analysis of thematic categories of Twitter terms highlighted that the most relevant portion of key terms consists in references to the time and location of the flood event. According to this result, Twitter might be primarily used as a means to disseminate warnings.

The analysis of Twitter profiles provided interesting insights on the most active users and the users that publish the most popular tweets.

The percentages of individual profiles among the most popular users (Fig. 5 a) let us suppose that Twitter followers prefer supporting individual accounts by liking their tweets, while they tend to retweet less frequently. Twitter followers probably prefer to retweet from official sources rather than individuals for a reliability reason.

Another inference can be made on the basis of the profile of the users who published at least ten tweets in one month (Fig. 5b). A small percentage (5%) of profiles that belong to the rescue service are among the most active users. This could be explained by the fact that rescue services usually centralise information management.

By looking closer at the popularity of the five most active users (Fig. S8), we could observe that, in this small sample, there is no correlation between the number of tweets and the number of likes or retweets. However, if the number of likes is high, the number of retweets will probably be high as well.

By focusing on the tweets that obtained more than fifty likes or more than fifty retweets, it is possible to observe that the most popular tweets are published by the press and public authorities, i.e. those actors that are also visible in the press. Percentages presented in Fig. 5 c and Fig. 5d suggest that Twitter users operating in the media sector and in the public administration/policy-making sector are leading opinion makers in the debate on Paris flood risk. The press seems to raise broad interest: in the changing landscape of digital media, the press continues to be considered as a source of reliable information. Public authorities and policy makers are frequently in the social media spotlight, as their views can have direct consequences for the society. Twitter is a media open to any contributor. However, it seems that the most popular users are those actors that are also visible in the press. Fig. 5 c and 5d also call attention to a widespread interest among Twitter users in the flood impact on transports and cultural heritage: the inhabitants of Paris region, as well as tourists and people travelling across the region, felt strongly affected by the flood impacts on museums and transport infrastructure. This is probably specific to a population that lives or travels in a metropolis with a dense transport network and a very high concentration of historical monuments and museums.

Lastly, the portion of users dealing with 'transports' is smaller in the case of the 'most retweeted users' (Fig. 5d) than in the case of the 'most liked users' (Fig. 5 c). A percentage difference of 4% that could be explained by the fact that these tweets describe how transport workers cope with the flood, but they don't convey helpful information for the passengers.

7 Conclusions and perspectives

rosa 18/3/19 17:48
Deleted: While network representation supported by Gargantext isn't suitable for a corpus of tweets, an

rosa 18/3/19 17:48
Deleted: is more appropriate for the purpose of this research.

rosa 18/3/19 17:50
Deleted: Moreover, Twitter data are valuable since they contain information on

rosa 18/3/19 17:52
Deleted: the

rosa 18/3/19 17:51
Deleted: of its users and metrics that describe how the users interact with each other. Indeed, it is possible to identify

rosa 18/3/19 17:51
Deleted:

rosa 18/3/19 19:26
Deleted: /

rosa 18/3/19 17:55
Deleted:

rosa 18/3/19 19:10
Deleted: , h

rosa 12/3/19 20:33
Deleted: this

rosa 12/3/19 18:57
Deleted: . ... [71]

In this research we employed big data exploration techniques to investigate how urban resilience to extreme weather is perceived in the digital media debate. Through this study, we firstly intended to test how these techniques can be used to define indicators of social representation of urban resilience. In our view, these indicators can be integrated to a wider assessment of urban resilience to weather extremes. Secondly, we aimed at gaining an insight of the social perception of flood resilience in two French urban areas. This research is still in progress, however, through the experiments presented in this paper, we obtained quantitative data on:

- The correlation existing between the intensity of the digital media debate (a social factor) and the level of the river discharge (an environmental factor);
- The evolution of the intensity of the debate over time, in two different locations and in two different media contexts (French press and Twitter);
- The differences that exist in terms of quality of the contents (i.e. reference to flood resilience solutions) between the press coverage of a flood event and the Twitter coverage of the same event;
- The topics and actors mentioned by the press that correspond to high degree nodes, and how these patterns change in two different urban areas;
- The most probable co-occurrences that exist among these topics and actors, and how these patterns change in two different urban areas;
- The most prominent topics in the Twitter debate;
- The profile of the most active and the most popular Twitter users.

The initial results are promising: these enable a complex understanding of the intensity and quality of the digital media debate. This research contributes to gain a better understanding of the public opinion, conveyed in the media, and the opinion makers which is beneficial for any urban resilience project in the Paris region. Indeed, this kind of analysis can contribute to creating a better connection with the urban community and optimise a project impact through dialogue and cooperation with the stakeholders.

In the future we intend to push forward our research by considering a longer time scale. Furthermore, this methodology can be easily applied to other urban areas, affected by different climate related stresses and shocks. We also intend to study the correlations that might exist between the intensity and quality of the debate and other resilience variables, such as: the number of citizens affected by extreme weather, the surface of revegetated areas, the amount of insurance compensation for natural disasters, etc. Quantitative analysis of the graphs could be further developed: a detailed analysis of other measures that characterise graphs would be necessary to confirm our hypothesis based on the analysis of the node degrees and edges weight. Tweets analysis could be more fruitful if supported with automated exploration of Twitter accounts and graph representation of likes and retweets. It would be then possible to analyse larger samples of data and easily move from an aggregated level to a detailed level of analysis.

Data availability. The supplement related to this article is available online at:

Acknowledgements. The authors are thankful for the technical support provided by

rosa 18/3/19 18:07
Deleted: article

rosa 13/3/19 17:32
Deleted: metrics

rosa 13/3/19 17:32
Deleted: metrics

rosa 18/3/19 10:04
Deleted: allowed us to obtain

rosa 18/3/19 18:08
Deleted: <#>The most central topics and actors in the press, and how these patterns change in two different urban areas;

rosa 18/3/19 10:30
Deleted: frequent

rosa 18/3/19 10:30
Deleted: connections

rosa 18/3/19 10:04
Deleted: se

rosa 18/3/19 10:04
Deleted: they allowed a

rosa 18/3/19 18:09
Deleted:

rosa 18/3/19 19:11
Deleted: regreened

rosa 18/3/19 19:10
Deleted: amount

rosa 18/3/19 19:11
Deleted: s

rosa 18/3/19 18:16
Deleted:

rosa 18/3/19 10:27
Deleted: network

rosa 13/3/19 17:38
Deleted:

rosa 18/3/19 18:22
Deleted: Supplements to this article are available online.

rosa 18/3/19 18:19
Deleted: We

Institut des Systèmes Complexes Paris Île-de-France and Frédérique Bordignon from École des Ponts ParisTech during the implementation of Europresse, Gargantext and Gephi. The authors gratefully acknowledge the financial support of the chair 'Hydrology for resilient cities' endowed by Veolia.

References

Boyd, A.: Broadcast Journalism, Techniques of Radio and TV News. Oxford: Focal, 1994.

Bruns, A. and Liang, Y.E.: Tools and methods for capturing Twitter data during natural disasters, First Monday, 17, 4, doi:10.5210/fm.v17i4.3937, 2012.

Bruns, A., Burgess, J., Crawford, K. and Shaw, F.: #qldfloods and @QPS Media: Crisis communication on Twitter in the 2011 south east Queensland floods, ARC Centre of Excellence for Creative Industries and Innovation, Brisbane, 58 pp., 2012.

Bucchi, M.: Style in science communication. Public Understanding of Science, 22, 8, 904–915, 2013.

Chacon-Hurtado, J. C., Alfonso, L. and Solomatine, D.: Dimensioning of precipitation citizen observatories in an uncertainty-aware context, EGU General Assembly, Vienna, Austria, 23–28 April 2017, EGU2017-18523-1, 2017.

Chavalarias, D.: Le Tweetoscope Climatique. Une représentation collective des enjeux autour du climat. La lettre de l'INSHS, 38, 39-42, 2015.

Chavalarias, D. and Delanöe, A.: Gargantext, available at: gargantext.org (last access: 18 May 2018.), 2017.

Climaps: Reading the state of climate change from digital media, available at: http://disq.us/t/1gj2hci (last access: 18 March 2019), 2013.

Cutter, S.L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E. and Webb, J.: A place – based model for understanding community resilience to natural disasters, Global Environ. Chang., 18, 598–606, 2008.

Cutter, S.L., Burton, C.G. and Emrich, C.T.: Disaster Resilience Indicators for Benchmarking Baseline Conditions, J. Homel. Secur. Emerg., 7, 1, 51, 2010.

Gaitan, S., Calderoni, L., Palmieri, P., Ten Veldhuis, M.C., Maio, D., and van Riemsdijk, M.B.: From Sensing to Action: Quick and Reliable

Access to Information in Cities Vulnerable to Heavy Rain. IEEE Senspors Journal, 14, 4175–4184, 2014.

rosa 18/3/19 18:20
Deleted: This research in led in the framework of the Chair "Hydrology for Resilient Cities" supported by Veolia

rosa 18/3/19 18:41
Moved up [15]: Blog du moderateur: blogdumoderateur.com, last access: 18 May 2018.

rosa 18/3/19 19:24
Deleted: -

rosa 18/3/19 19:24
Deleted: -

rosa 18/3/19 18:41
Deleted:
Gargantext: gargantext.org

rosa 18/3/19 18:44
Moved (insertion) [16]

rosa 18/3/19 18:44
Deleted: Gargantext: gargantext.org

rosa 18/3/19 18:44
Deleted: (D. C

rosa 18/3/19 18:44
Deleted: A.

rosa 18/3/19 18:45
Deleted: 2017),

rosa 18/3/19 18:44
Deleted:

rosa 18/3/19 19:26
Deleted: –

rosa 18/3/19 18:42
Deleted: Europresse: europresse.com, last access: 18 May 2018.

rosa 18/3/19 19:24
Deleted: -

Houston, J.B., Pfefferhaum, B. and Rosenholtz, C.E.: Framing and Frame Changing in Coverage of Major U.S. Natural Disasters, 2000–2010, Journalism and Mass Communication Quarterly, 89, 4, 606–623, 2012.

Keating, A., Campbell, K., Mechler, R., Michel Kerjan, E., Mochizuki, J., Kunreuther, H., Bayer, J., Hanger, S., McCallum, I., See, L., Williges, K., Atreya, A., Botzen, W., Collier, B., Czajkowski, J., Hochrainer, S. and Egan, C.: Operationalizing Resilience Against Natural Disaster Risk: Opportunities, Barriers and A Way Forward, Zurich Flood Resilience Alliance, 43 pp, 2014.

Lanfranchi, V., Ireson N., When U., Wrigley S.N. and Fabio C.: Citizens' Observatories for Situation Awareness in Flooding, in: Hiltz, S.R., Pfaff, M.S., Plotnick, L. and Shih P.C. (ed.) Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2014): 18–21 May 2014, University Park, Pennsylvania, USA, 145–154, 2014.

Latour, B.: The whole is always smaller than its parts. A digital test of Gabriel Tarde's Monads, British Journal of Sociology, 63, 4, 591–615, 2012.

Leydesdorff, L. and Milojevic, S.: Scientometrics, in: International Encyclopedia of Social and Behavioural Sciences, Section 8.5: Science and Technology Studies, Subsection 85, 030, Lynch, M. (ed), Elsevier, 2015.

Mangalagiu, D., Wilkinson, A. and Kupers, R.: When futures lock-in the present: Towards a new generation of climate scenarios, in: K. Hasselmann, C. and Jaeger, C. (ed.) Reframing the Problem of Climate Change: From Zero Sum Game to Win-Win Solutions, Routledge, 2012.

Morss, R., Demuth, J., Lazrus, H., Palen, L., Barton, C. Davis, C. Snyder, C., Wilhelmi, O., Anderson, K., Ahijevych, D., Anderson, J., Bica, M., Fossell, K., Henderson, J., Kogan, M., Stowe, K. and Watts, J.: Hazardous Weather Prediction and Communication in the Modern Information Environment, Bull. Amer. Meteor. Soc., doi:10.1175/BAMS-D-16-0058.1, 2017.

Newman, M. E. J. Networks: an Introduction. Oxford: Oxford UP, 2010.

Niederer, S.: Global warming is not a crisis! : Studying climate change scepticism on the Web. NECSUS European Journal of Media Studies, 2, 83–112, DOI:10.5117/NECSUS2013.1.NIED, 2013.

OECD: Seine Basin, Île-de-France, 2014: Resilience to Major Floods, OECD Reviews of Risk Management Policies, Éditions OCDE, Paris, 204 pp., 2014.

Palen, L., Starbird, K., Vieweg, S. and Hughes, A.: Twitter–based information distribution during the 2009 Red River Valley flood Threat, Bul. Am. Soc. Info. Sci. Tech., 36, 13–17, doi:10.1002/bult.2010.1720360505, 2010.

Resilience Alliance: Assessing resilience in social–ecological systems: Workbook for

rosa 18/3/19 18:44
Moved up [16]: Gargantext: gargantext.org (D. Chavalarias and A. Delanöe, 2017), last access: 18 May 2018.

rosa 18/3/19 18:46
Deleted: Gargantext: gargantext.org (D. (D. Chavalarias and A. Delanöe, 2017), last access: 18 May 2018.Gephi: gephi.org ... [72]

rosa 18/3/19 19:24
Deleted: -

rosa 18/3/19 19:26
Deleted:

rosa 18/3/19 19:24
Deleted: -

rosa 18/3/19 19:24
Deleted: -

rosa 18/3/19 19:12
Deleted: k

rosa 18/3/19 19:26
Deleted: –

rosa 18/3/19 19:24
Deleted: -

rosa 18/3/19 19:26
Deleted: –

practitioners. Version 2.0., Resilience Alliance, 54 pp., 2010.

Rogers, R. and Marres N.: Landscaping climate change: A mapping technique for understanding science and technology debates on the World Wide Web. Public Understanding of Science, 9, 141–163, DOI:10.1088/0963- 6625/9/2/304, 2000.

Trench, B.: Internet: turning science communication inside-out?, in: Bucchi, M. and Trench, B. (ed.) Handbook of Public Communication of Science and Technology. Rutledge, 2008,

UN/ISDR: Indicators of Progress: Guidance on Measuring the Reduction of Disaster Risks and the Implementation of the Hyogo Framework for Action, United Nations secretariat of the International Strategy for Disaster Reduction (UN/ISDR), Geneva, Switzerland, 59 pp., 2008.

Venturini T., Baya Laffite, N., Cointet, J.P., Gray, I., Zabban, V., De Pryck, K.: Three maps and three misunderstandings: A digital mapping of climate diplomacy, Big Data and Society, 1, 2, 1–19, 2014.

Vieweg, S., Hughes, A. L., Starbird, K., and Palen, L.: Microblogging during two natural hazard events: what Twitter may contribute to situational awareness, in: CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, USA, 10–15 April 2010, 1079–1088, 2010.

rosa 18/3/19 19:12
Deleted: outld

rosa 18/3/19 18:49
Deleted: ... [73]

rosa 18/3/19 19:24
Deleted: -

rosa 18/3/19 19:24
Deleted: -

rosa 18/3/19 18:51
Deleted: ... [74]

Supplement S1: Details on how Gargantext works and step-by-step implementation.

With Gargantext it is possible to extract, automatically as well as manually, a list of key terms from a corpus of digital texts. A first list of terms is automatically extracted by Gargantext algorithms on the basis of their occurrence, compared to other occurrences that characterise all Gargantext database corpora, as well as on the basis of the co-occurrences that characterise the specific corpus. This list can be manually modified: enriched with other terms extracted from the corpus, reduced, some terms can be merged in sub-lists. Gargantext is unique in terms of ergonomics. Indeed, it is possible to analyse of texts at different levels: a micro level (the manual selection of key terms in a single document), a meso level (the manual selection of key terms and creation of groups of terms in a table of terms extracted by Gargantext), a macro level (the manual selection of key terms in the graph representation).

The graph representations presented in this research are computed on the basis of the semantic proximity measure between pairs of key terms (from the list) that is called 'conditional distance'. Co-occurrence graphs based on conditional distance illustrate which terms co-occur in the same meaning unit (e.g. a press article) and with what probability. As it is specified in Gargantext documentation for advanced users (iscpif.fr/gargantext/mesures-utilisees-dans-gargantext/, last access: 18 March 2019), 'the conditional measure P_c between term *i* and term *j* (...) is the maximum of the two conditional probabilities between *i* and *j*. If n_i (n_j respectively) is the number of occurrences of *i* (*j* respectively) in the corpus of articles and n_{ij} is the number of co-occurrences, we will have the following formula":

$$P_c = max\left(\frac{n_{ij}}{n_i}, \frac{n_{ij}}{n_j}\right)$$

Gargantext computes non-directed graph representations, i.e. these graph have no directed edges. The graph representations remain stable even when few documents are deleted from the corpus or few nodes are removed from the graph. Gargantext graphs are weighted a "weight" from zero to one is assigned to each edge of the graph and it indicates the probability that two terms co-occur. The degree of each node (i.e. the number of edges connected to the node) can be displayed through Gargantext graph visualisation engine.

The nodes are assembled in cohesive subsets through a clustering algorithm, more specifically through the Louvain modularity. The Louvain method for community detection is used to maximise the graph modularity. A graph 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 graph 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 presented in Gargantext documentation (iscpif.fr/gargantext/, last access: 18 March 2019).

This research was supported by Gargantext, but it required the following manual interventions that were carried out by the authors:

1) We selected and downloaded the two corpora of press articles (articles covering the 2016 Seine river flood and the articles covering the 2015 Alpes-Maritimes flood) from Europresse in HTML format;
2) We uploaded the file on Gargantext;
3) We selected, among the terms automatically extracted by Gargantext from the two corpora, the most relevant terms and we merged synonyms, declensions of terms and equivalent forms;
4) We kept only those terms that occur at least five times in one of the corpora;
5) We verified the daily distribution of terms referring to the resilience solutions (through Gargantext Analytics view) and we recorded the daily incidence values in an Excel file in order to generate a histogram (Fig. 1s);
6) We added to the key term list the terms referring to actors and affected infrastructure (with at least five occurrences) in order to represent in the graph the stakeholders involved in flood risk management and the context where the resilience solutions were implemented;
7) We launched on Gargantext the graph representation based on conditional distance between pairs of terms from the term list;

8) We launched, through the graph visualisation engine, the Louvain algorithm in order to position the strongly related nodes close to each other;
9) We selected the Gargantext option that displays the terms corresponding to each node and the node degrees;
10) We zoomed in the graph to observe all the nodes (even those with a small degree) and we captured a photo of the two graphs;
11) We extracted the graph in. gexf format in order to analyse it with Gephi software;
12) We imported the. gexf file in Gephi and we converted it in two Excel tables with the node degrees and the edge weights;
13) We generated through Excel the figures presented in the Supplements.

Table S2.1: Key terms related to flood resilient solutions (in French). The terms were extracted from the articles on the 2016 Seine River flood and the 2015 Alpes-Maritimes flood. The list was automatically created by Gargantext algorithms, then it was manually refined on the basis of the relevance of the terms.

Adaptation, adapter, adapté, adaptations, adaptées, alerte, alertes, alerte générale crue, alerte orange, alertes orange, vigilance orange, alerter, alerte rouge, anticiper, assurance, assuré, assurée, assurés, assurances, barrages, barrage, barrer, barrer l'accès, barrières, barrières étanches, barrière, barrières en inox, aquabarrières, aquabarrière, aqua-barrière, aqua-barrières, barrières anti-crue, bassin de rétention, bassins de rétention, bassins d'orage, casier, expérimental, casiers, réservoir, casier, réservoirs, réservoir, casier-pilote, batardeaux, batardeau, bulletin, bulletins, capteurs, stations de mesures, courantomètres hydroacoustiques, équipements de mesure, courantomètres hydro-acoustiques, station hydrométrique, capteur, cellule de crise, centrale de crise, constat, constats, coupés, coupez, coupées, coupée, barrées, coupé, culture du risqué, débat public, dépanneuses, dépanneuse, désimperméabilisation, désimperméabilisations, végétalisation, diagnostics, diagnostic, digues, digue, données, dons, écluse, écluses, école fermée, écoles fermées, entretien des cours d'eau, espaces verts, espace vert, estimations, estimation, état de catastrophe naturelle, cat nat, plan d'aide, solidarité financière, dispositifs de soutien, classement en catastrophe naturelle, fonds d'indemnisation des catastrophes naturelles, garantie de catastrophe naturelle, procédure de catastrophe naturelle, système de solidarité, fonds de solidarité, étude, études, évacuation, évacuations, évacuées, évacuer, évacuation préventive, évacuations préventives, évaluent, évaluation, évalue, évaluations, fonds exceptionnel de soutien, former, gestion des milieux aquatiques et prévention des inondations, hébergement, hébergées, centre d'hébergement d'urgence, centre d'accueil, centres d'hébergement d'urgence, refuge, héberger, relogées, relogés, relogée, relogé, structures d'accueil, héberge, hébergements, relogement, information, informés, renseignements, informée, informer, informations, informé, interdiction, interdite, interdictions, interdit, interrompu, suspension totale, arrêtés, interrompue, arête, interrompus, arrêté, la bassée, bassée, lacs réservoirs, lacs-réservoirs, lacs de rétention, lacs de stockage, lac-réservoir, retenues des grands lacs, lacréservoir, lacs de retenue, lac réservoir, lacs artificiels, quatre grands lacs, lacs-réservoir, lidar, mémoire, mémoires, mesure de sécurité, consignes de sécurité, règles générales de sécurité, mesures préventives, mesure préventive, mettre en sécurité, mis en sécurité, mise en sécurité, modélisation, modèles, modéliser, murer, obturer, muret, murettes, murets, nouvelles zones endiguées, nouvel ouvrage, oeuvres stockées, parpaings, plan communal, plan communal de sauvegarde, pcs, plans communaux, plans communaux de sauvegarde, plan crue, plans de prévention, plans de prévention des risques, plan de prévention, dispositifs de prévention, plans de préventions des risques naturels, plans de préventions du risque d'inondations, ppr, pprn, stratégies de prévention, programmes d'actions de prévention des inondations, plans de prévention du risque inondation, ppri, plans de protection, plan de protection, plaques, pompes, pompes de relevage, pompage, pompe, pompant, possibles évacuations, précaution, précautions, première zone pilote, préparation, préparer, préparatifs, préparés, prévenir, prévention, préventive, préventives, titre préventif, prévisions, prévoit, prévision, prévoir, prise de conscience, prendre conscience, conscientiser, conscientisation, protection, protections, recherche, recherches, reconstruction, régulation, régulateur, relevés, relevé, relevés automatiques, mesures de débit, renforcer, renforcé, renfort, renforts, réparations, reconstruire, réhabilitation, réparation, repères, repère, restrictions, restriction, retour d'expérience, sauvetage, sauvetages, scénario, secouru, secourue, sécuriser, sensibilisation, réveillé les consciences, sensibilisations, conscience du risque, sensibilisant, service de prévention vigicrues, simulation, simulateur, simulations, site vigiecrues, solidarité, solidaires, solidarités, solution technique, soutien financier, financer, aides financières, fonds, rembourser, indemnisés, indemniser, fonds d'urgence, subventions, dédommagements, remboursements, indemnisations, moyens financiers, fond, financement, indemnisation, remboursement, dédommager, indemnisé, crédit, aide financière, subvention, financements, finances, indemnisées, stocker, stockages, stockage, stockant, suivi, surveillance, surveillances, surveiller, surveille, surveillée, vanne, vannes, vigicrues, vigicrue, vigiecrues, vigilance jaune, vigilance rouge, volets en acier

Table S2.2: English translation of the key terms related to flood resilient solutions presented in Tab. S2.

Adaptation, adapt, adapted, adaptations, adapted, alert, alerts, general flood alert, orange alert, orange alerts, orange vigilance, to alert, red alert, anticipate, insurance, insured, insurance, dams, dam, to bar, to block the access, gates, barriers, barrier, waterproof barriers, water barrier, water-barriers, water barrier, water barriers, stainless steel barriers, flood prevention barriers, storm water basin, storm water basins, locker, experimental, basins, retention reservoir, basin, retention reservoirs, pilot basin, cofferdams, cofferdam, bulletin, bulletins, sensors, measuring stations, hydroacoustics current meters, measuring equipment, hydro-acoustics current meters, hydrometric station, sensor, crisis cell, crisis centre, report, findings, cut, barred, risk culture, public debate, recovering trucks, recovering truck, increase permeable surfaces, revegetation, diagnoses, diagnosis, dikes, dyke, data, donations, lock, locks, closed school, closed schools, maintenance of streams, green spaces, green space, estimates, estimate, state of natural disaster, cat nat, aid plan, financial solidarity, support systems, natural disaster classification, natural disaster compensation fund, natural catastrophe guarantee, natural disaster procedure, solidarity system, solidarity fund, study, studies, evacuation, evacuations, preventive evacuation, assess, evaluation, assesses, evaluations, exceptional support fund, to train, management of aquatic environments and flood prevention, accommodation, shelter, emergency shelter, emergency shelters, shelter, relocated, to relocate, hosted, sheltering structures, hosts, accommodation, relocation, information, informed, to inform, prohibition, prohibited, prohibitions, forbidden, interrupted, total suspension, stopped, interrupted, ridge, interrupted, arrested, bassée, reservoir lakes, reservoir lakes, retention lakes, artificial lakes, four large lakes, lidar, memory, memories, safety measure, safety instructions, general safety rules, preventive measures, preventive measure, safety, to secure, secured, modelling, models, modelling, to wall up, to obstruct, low wall, low walls, new dyked areas, new construction, stored art pieces, concrete blocks, municipality plan, municipality plan of safeguard, municipality plans, municipality plans of safeguard, flood plan, prevention plans, risk prevention plans, prevention plan, preventive measures, natural risk prevention plans, flood risk prevention plans, prevention strategies, flood prevention action programs, flood risk prevention plans, protection plans, protection plan, plates, pumps, recovery pumps, to pump, pump, pumping, possible evacuation, precaution, precautions, first pilot area, preparation, prepare, prepared, prevent, prevention, preventive, preventives, precautionary, forecasts, forecast, predict, awareness, protection, protections, research, reconstruction, regulation, regulator, survey, automatic surveys, flow measurements, reinforce, reinforced, reinforcement, reinforcements, reparations, rebuild, rehabilitation, reparation, landmarks, landmark, restrictions, restriction, feedback, rescue, rescues, scenario, rescued, secure, awareness raising, awakened consciences, awareness of risk, Vigicrues prevention service, simulation, simulator, simulations, Vigiecrues site, solidarity, solidarities, technical solution, financial support, finance, financial aid, funds, repay, indemnified, compensate, emergency funds, grants, compensation payments, reimbursements, indemnifications, financial means, funds, financing, reimbursement, compensation, credit, financial assistance, grant, financing, finances, indemnified, storage, stocking, monitoring, surveillance, monitor, monitors, monitored, valve, valves, yellow vigilance, red vigilance red, steel shutters.

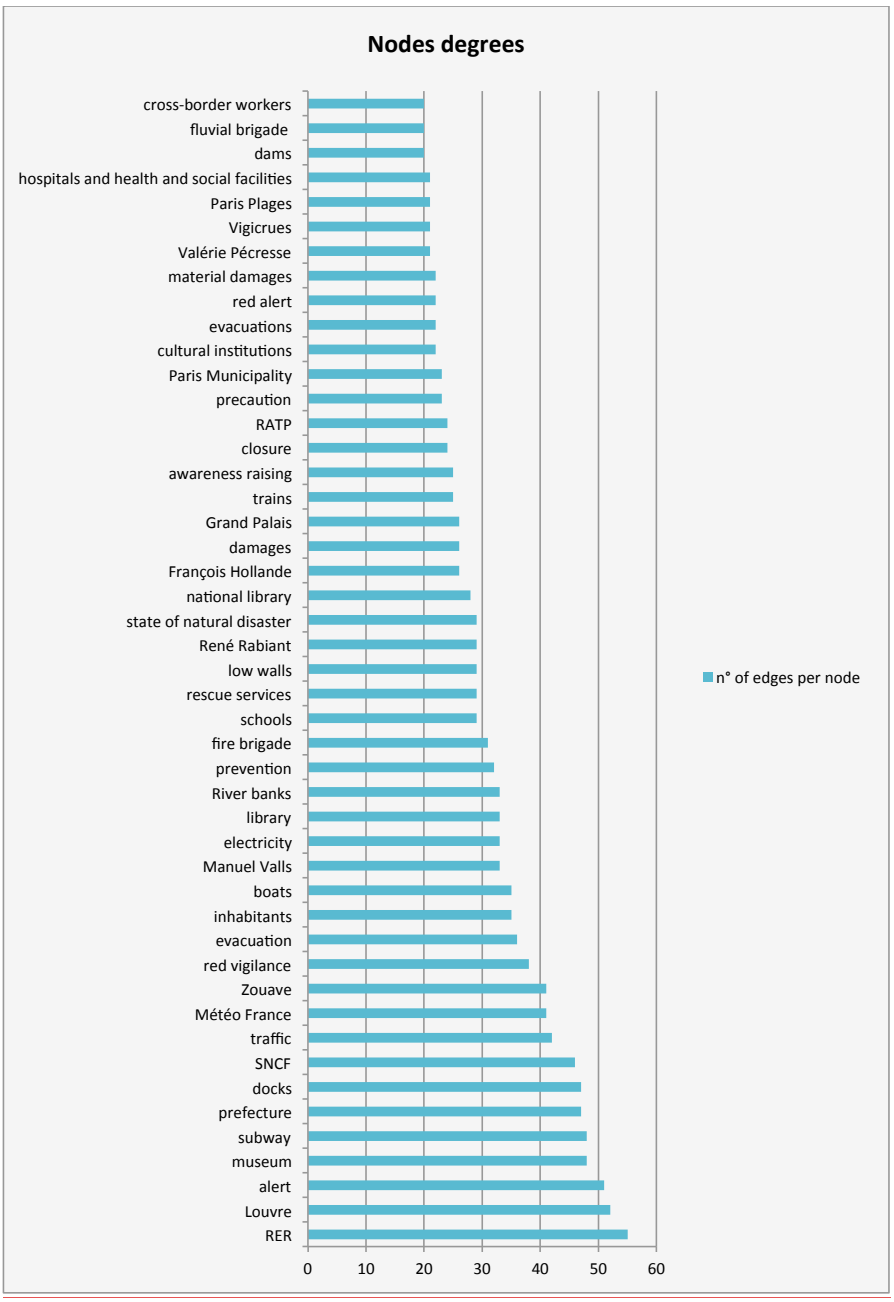
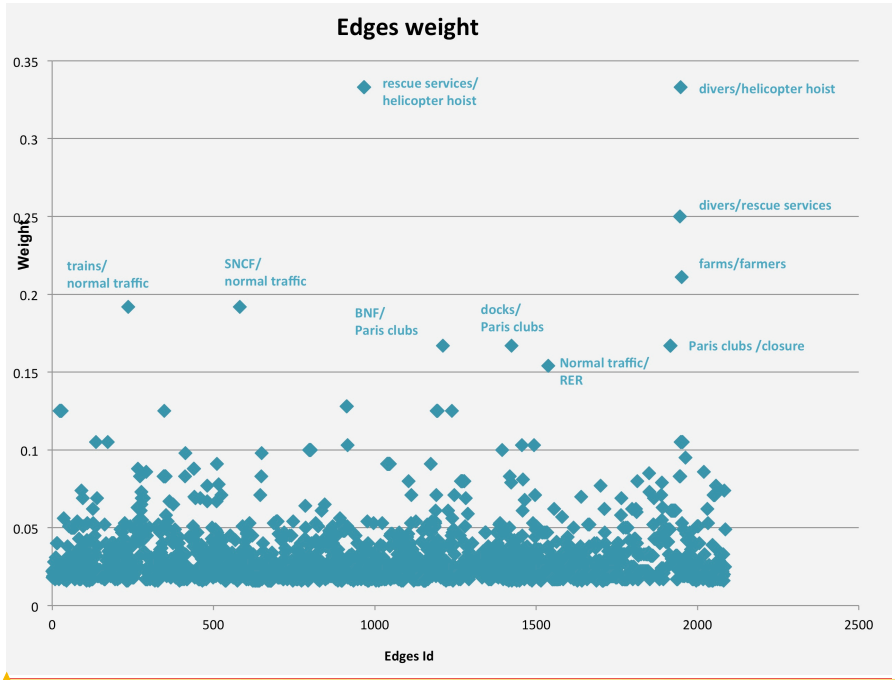


Figure S3.1: The key terms corresponding to the nodes with the highest degree in the graph. The graph representation was computed on the basis of the measure of conditional proximity between pairs of key terms in the corpus of press articles on the Seine River flood.



Unknown

Formatted: Font:(Default) Times New Roman, 9 pt, Bold

Figure S3.2: The most probable term co-occurrences corresponding to the edges with the highest weight in the graph. The graph is based on the corpus of press articles on the 2016 Seine River flood.

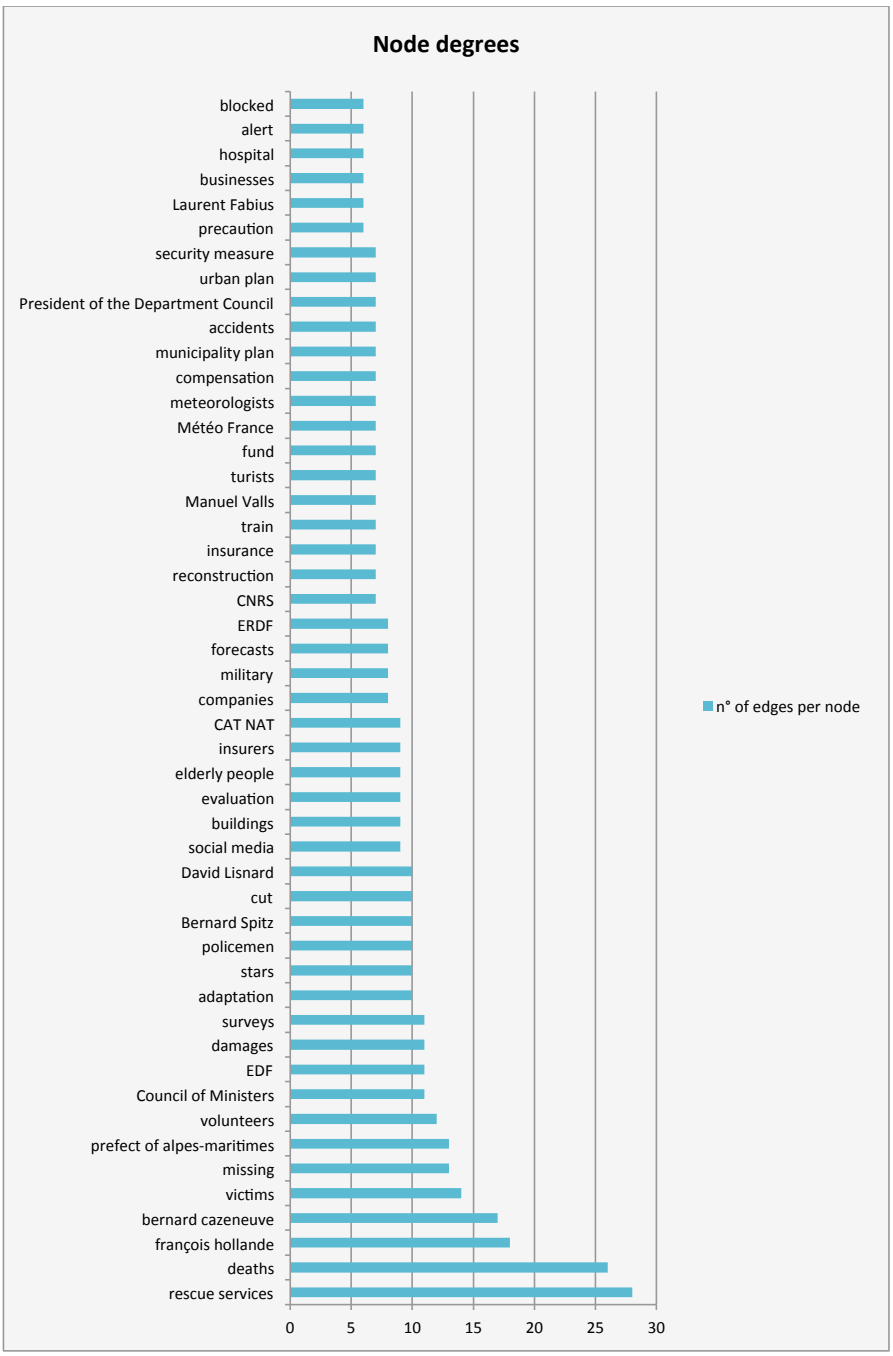


Fig. S4.1: The key terms corresponding to the nodes with the highest degrees in the graph. The graph representation is based on a corpus of press articles on the 2015 Alpes-Maritimes flood: it was computed on the basis of the measure of conditional proximity between pairs of key terms in the corpus of articles.

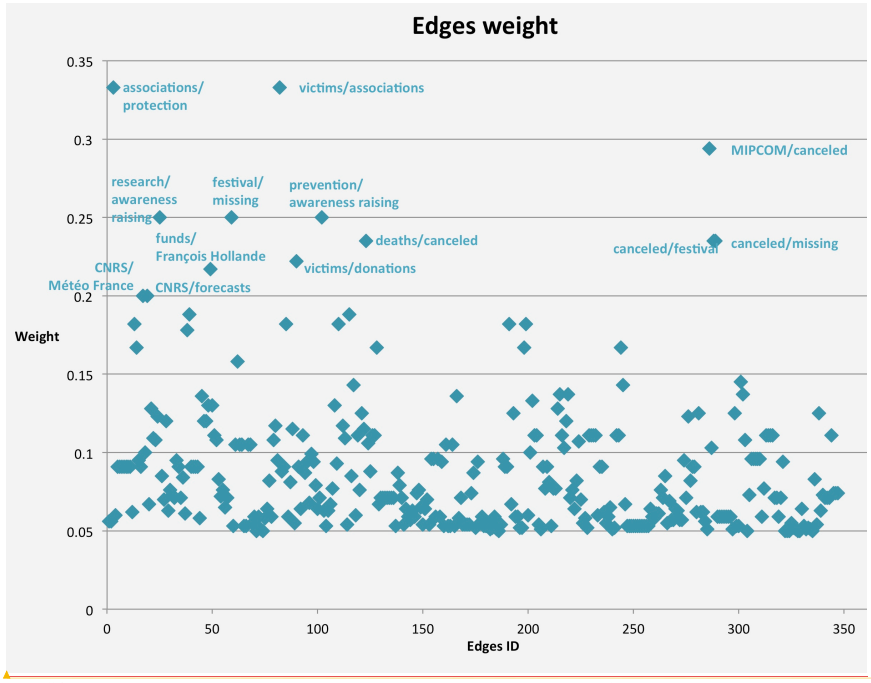


Fig. S4.2: The most probable term key term co-occurrences corresponding with the highest weight in the graph. The graph is based on a corpus of press articles on the 2015 Alpes-Maritimes.

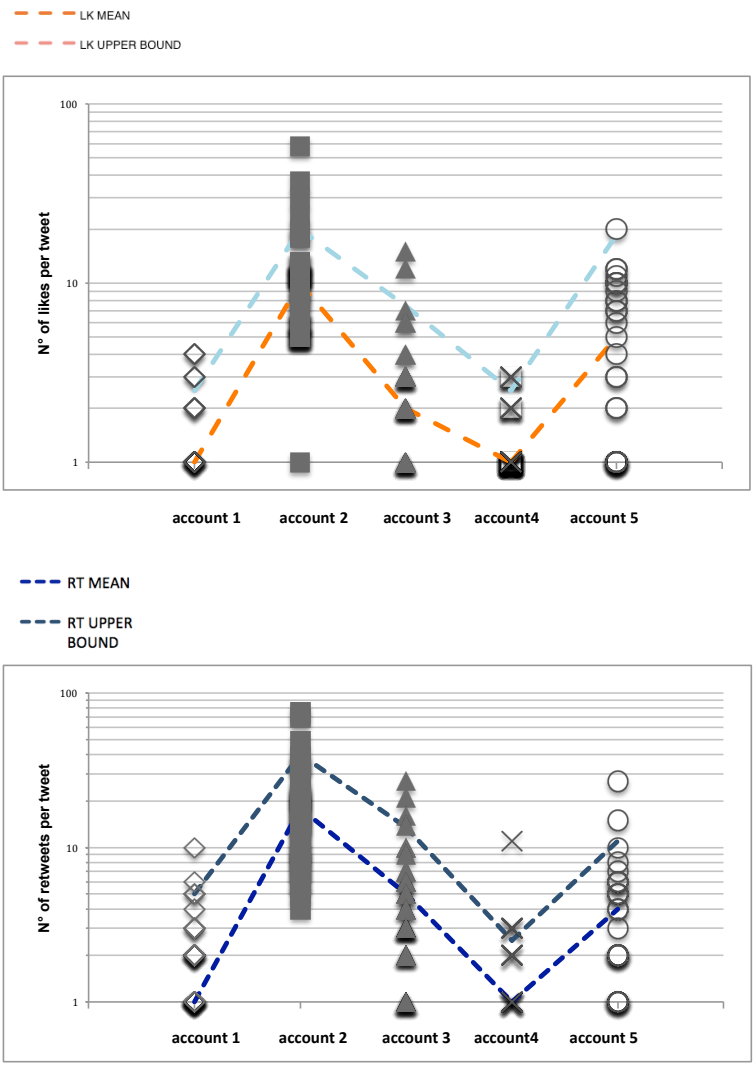
Unknown

Formatted: Font:(Default) Times New Roman, 9 pt, Bold, Font color: Custom Color(RGB(106;106;106))

Table S5.1: Thematic groups of key terms based on the tweets on the 2016 Seine River flood. The key terms were extracted from the tweets corpus (in French).

Thematic group (or category)	*Key terms extracted from the corpus of tweets*
FLOOD RESILIENCE SOLUTIONS	agents, opération, mobilisés, exercices, nous intervenons, urgent, une aide, soutien, barrage, catastrophenaturelle, cellules de crise, assurances, vigilances, vigilance orange, vigicrue, alert, aides, vigilance rouge, pompiers, vigilance jaune, fonds, police, prévention, sécurité, alerte rouge, secours, surveillance, appel aux dons, évacuation, sont mobilisées, alerte orange, vigilance orange, plouf, urgent, simulation, nous intervenons, catnat, précaution, bénévoles, sauvetage, mobilisation, gemapi, réparations, bétons, pcsorsay, catastrophesnaturelles, reconnaissance, travaux, mission, watergate, culture, aloa
MEDIA CONTENT / SOURCE	photos, periscope, info, photography, images, direct sur, vidéos, photographe, pic, view, news, artwork, streetphotography, cartes, journal, pictures, photooftheday, nofilter, picoftheday, faitsdivers, longexposure @ paris, photography https, latergram, une image, infographie, actu, communiqué, msgu, video, blog, dessin, qag, rediff, hashtag, www.youtube.com/watch, sky, flickr, afp, nouvellerepublique, bfmtv, france2, septahuit, facebook
TIME/LOCATION	matin, cet après-midi, week, hier soir, lundi, yesterday, tomorrow, Sunday, spring, départements, nemours, communes, essonne, city, alma, juvisy, yvelines, idf, bercy, europe, montargis, marne, melun, ville, poissy, régions, chatou, austerlitz, puteaux, valdemarne, pontneuf, boulognebillancourt, seine-et-marne, longjumeau, hautsdeseine, pont mirabeau, bièvre, courbevoie, pont alexandre iii, issy, cité, jatte, neuilly, eiffel, bastille, iles, concorde, village, palaisdejustice, sites, paname, quaideseine, yerres, bougival, ile saint louis, iledelacite, grigny, campagne, louvre, pontdelalma, suresnes, georges pompidou, reuil, pontmirabeau, alfortville, asnieres, notredame, cathedral, notre-dame, notre dame, paris plage, orsay, tunnel, château, park, lyon, statueofliberty, sur berges, arts
RISK	risque, zone inondable, crue centennale, danger, crue historique, crise est hautement probable, menace, threat, soyez prudent, soyez vigilant
IMPACT	touché, catastrophe, cause, boue, stations, millions d'euros, milliards d'euros, inondée, charges, crise, streets, ports, patrimoine, cars, bateaux, sculptures, difficulté, navigation, débordé, débordements, bilan, sécher, rouvertes, impacts, laseinedeborde, major flood, coupé, bloqué, bridge, sinistrées, circulation, museum, banks, dégâts, tourisme, boats, cultures, victims, rerc, fermées, pollutions, œuvre, trains, maison, mort, agriculture, rue, archives, voitures, stations, transports, collection, bouchons, expo, facture, cars, metro, sinistres, gare, vache, inondés, pib, concert, assainissement, batobus, immobilier, moustique, restaurant
CAUSES	climatechange, climate, chgt climatique, sous-estimée, sous-estimé, développement urbain
WEATHER EVENTS	décrue, highest level, météo, rain, meteo, record, mesure, hauteurs, unfetter, water levels, level, monte, seine baisse, metres, decrue, feet, seine level, weather, stabilisation, landmarks, comparaison, meters, ça monte, grêle, averses, déluge, normal levels, niveau, précipitations, zouave, nasa, evolution, cumul
OTHER NEWS	loitravail, greve, grevessncf, nonddl, grèves, attentats, brexit, manif, migrants, cgt, nuitdebout
EMOTIONAL CONTENT/ HUMOUR	parisjetaime, thankful, love, courage, impressionant, fluctuatnecmergitur, hope, pariseloveyou, parismaville, merci à, une pensée, bottes, humour, swim, uber, insolite, pigeon, venise,
STAKEHOLDERS	hidalgo, particuliers, habitants, conseil, masson delmotte, personnes, mairie, maires, populations, people, particuliers, collectivités, communes concernées, maire, vlacroute, slefoll, grandparis, seveso, conseil des ministres, comité de bassin, sncf, parisiens, hollande, valls, enfants, ratp, entreprises, fraeco, ccr, courtier, clients, tourist, experts, tpe, AFA, IPRG, pme, agriculteurs, segolene

Table S5.2: Thematic groups of key terms based on the tweets on the 2016 Seine River flood. The key terms were extracted from the tweets corpus. (English translation)	
Thematic group (or category)	***Key terms extracted from the corpus of tweets***
FLOOD RESILIENCE SOLUTIONS	agents, operation, mobilised, exercises, we intervene, urgent, help, support, dam, naturaldisaster, crisis cells, insurance, vigilance, orange vigilance, Vigicrue, alert, red vigilance, firefighters, yellow vigilance, funds, police, prevention, security, red alert, rescue services, surveillance, call for donations, evacuation, are mobilised, orange alert, orange vigilance, PLOUF, urgent, simulation, CATNAT, precaution, volunteers, rescue, mobilisation, GEMAPI, reparations, concrete, PCSOrsay, recognition, works, mission, watergate, culture, aloa
MEDIA CONTENT / SOURCE	photos, periscope, information, photography, images, direct, videos, photographer, pic, view, news, artwork, streetphotography, maps, newspaper, pictures, photooftheday, nofilter, picoftheday, longexposure @ paris, photography https, latergram, image, infographic, news, press release, social media in emergency management, video, blog, dessin, qag, rediff, hashtag, www.youtube.com/watch, sky, Flickr, AFP, NouvelleRepublique, BFMTV, FRANCE2, SEPTAHUIT, Facebook
TIME/LOCATION	morning, this afternoon, last night, Monday, yesterday, tomorrow, Sunday, spring, departments, Nemours, municipalities, Essonne, city, Alma, Juvisy, Yvelines, Ile-de-France, Bercy, Europe, Montargis, Marne, Melun, city, Poissy, regions, Chatou, Austerlitz, Puteaux, ValdeMarne, Pontneuf, BoulogneBillancourt, Seine-et-Marne, Longjumeau, HautsdeSeine, Mirabeau bridge, Bièvre, Courbevoie, Alexandre III bridge, Issy, Cité, Jatte, Neuilly, Eiffel, Bastille, islands, Concorde, village, courthouse, sites, Paname, SeineRiverbnks, Yerres, Bougival, Saint Louis Island, ÎledelaCité, Grigny, countryside, Louvre, Almabridge, Suresnes, Georges Pompidou, Reuil, Mirabeau bridge, Alfortville, Asnieres, NotreDame, cathedral, Notre-Dame, Notre Dame, Paris Plage, Orsay, tunnel, castle, park, Lyon, statueofliberty, on the banks, arts
RISK	risk, flood-prone area, one-hundred-year flood, danger, historic flood, highly probable crisis, threat, be prudent, be careful
IMPACT	hit, disaster, cause, mud, stations, million euros, billion euros, flooded, charges, crisis, streets, ports, heritage, cars, boats, sculptures, difficulty, navigation, overflowed, overflows, evaluation of losses, dry, reopened, impacts, theSeineRiveroverflows, major flood, cut, blocked, bridge, disaster stricken population, traffic, museum, banks, damage, tourism, cultures, victims, RERC, closed, pollution, artwork, trains, house, death, agriculture, street, archive, cars, stations, transports, collection, traffic jam, expo, bill, cars, subway, accidents, station, cow, flooded, GDP, concert, waste water system, batobus, real estate, mosquito, restaurant
CAUSES	climatechange, climate, underestimated, urban development
WEATHER EVENTS	decrease of water levels, highest level, weather, rain, weather, record, measure, heights, unfetter, water levels, level, rise, seine decreases, meters, decreases, feet, seine level, weather, stabilisation, landmarks, comparison, meters, it goes up, hail, showers, deluge, normal levels, level, precipitations, Zouave, NASA, evolution, cumulation
OTHER NEWS	labourlaw, strike, SNCFstrike, notoDDL, strikes, terrorist attacks, Brexit, protests, migrants, CGT, nuitdebout
EMOTIONAL CONTENT/HUMOUR	parisIloveyou, thankful, love, braveness, impressive, fluctuatnecmergitur, hope, pariseloveyou, parismycity, thanks to, a thought for, boots, humour, swim, uber, unusual, pigeon, Venice,
STAKEHOLDERS	Hidalgo, individuals, inhabitants, council, Masson Delmotte, persons, City Hall, municipalities, populations, people, local authorities, affected municipalities, mayor, VLacroute, SIeFoll, GrandParis, Seveso, Council of Ministers, Basin Comity, SNCF, Parisians, Hollande, Valls, children, RATP, companies, FRAECO, CCR, Courtier, clients, tourist, experts, TPE, AFA, IPRG, SMEs, farmers, Segolene



Figures S6: The users' behaviour: number of likes and number of retweets received by the five most active users for each of their tweets. These data were extracted from the sample of tweets on the 2016 Seine River flood.