Hybrid Crowdsensing: A Novel Paradigm to Combine the Strengths of Opportunistic and Participatory Crowdsensing

Marco Avvenuti †
marco.avvenuti@unipi.it

Salvatore Bellomo ‡
s.bellomo@iit.cnr.it

Stefano Cresci †‡
s.cresci@iit.cnr.it

Mariantonietta N. La Polla ‡
m.lapolla@iit.cnr.it

Maurizio Tesconi ‡
m.tesconi@iit.cnr.it

† Department of Information Engineering, University of Pisa, Italy
‡ Institute for Informatics and Telematics, National Research Council, Italy

ABSTRACT

Crowdsensing systems can be either participatory or opportunistic, depending on whether the user intentionally contributes data, or she simply acts as the bearer of a sensing device from which data is transparently collected. In this paper, we propose *hybrid* crowdsensing, a social media-based paradigm which aims at combining the strengths of both participatory and opportunistic crowdsensing. With hybrid crowdsensing, possibly relevant data is collected via an opportunistic approach. Then, users that spontaneously contributed are directly contacted and asked to provide additional information following a participatory approach. To demonstrate its feasibility and usefulness, we experimented the proposed paradigm for involving Twitter users in an emergency relief scenario. For each of the two real-world experiments we analyze the answer ratio to our questions, their time distribution, and responders' willingness to collaborate. Results support the adoption of hybrid crowdsensing, especially in those practical scenarios where users are emotionally involved.

Keywords

Crowdsensing; collective intelligence; online question answering; Twitter.

1. INTRODUCTION

In the last decade, the rapid growth of social networking platforms and the ubiquitous proliferation of mobile devices produced a great interest in studying how massive real-time social data can be used as a mine of information in domains such as health, transportation, energy, smart cities, intelligence and social/political crisis [29]. Within this context, a sensor is not only a physical device – as assumed in mobile crowdsensing – but also a logical or social metaphor implemented by the "human as a sensor" paradigm [2]. In literature, this paradigm is also referred to as "social sensing", "citizen sensing", or "crowdsensing", giving focus to the involvement of a number of people [1].

Because of their massive number of users, real-time features and ease-of-use, social networking platforms such as Twitter and Facebook have been the source of information for many crowdsensing systems. Depending on their awareness and their involvement in the system, human sensors are faced with a two-fold approach [17]: with *opportunistic* sensing, users spontaneously collect and share data as they go for their daily life, and relevant data is then transparently intercepted by a situation-aware system [16]; in contrast, with *participatory* sensing users consciously opt to meet an application request out of their own will, e.g., by photographing locations or discussing events and by intentionally sending such information to the sensing system.

Systems exploiting participatory sensing must usually provide some incentive to the users to perform the sensing action [32]. Thus, a key challenge in participatory sensing is the attraction of a significant user base. This may pose serious limitations to newly deployed systems and may ultimately lead to unsatisfactory results due to the lack of sufficient data. Another limitation of participatory systems may lie in the rigid way users provide their contributions (e.g., by filling a Web survey or adhering to a predefined format). These issues have been partly addressed within the field of online question answering (Q&A), where only a small subset of "expert" users are asked to provide contributions, with little restrictions [33, 24]. On the other hand, even if opportunistic sensing platforms do not require a specific user base, since they rely on already publicly available data, here the challenge is posed by the acquisition, preprocessing, and analysis of unstructured and heterogeneous data (e.g., tweets or Facebook posts) that is not specifically targeted to the sensing system [4, 3].

Contributions. Current crowdsensing approaches have both strengths and downsides. In this paper, we propose *hybrid crowdsensing*, a novel approach able to combine the strengths of participatory and opportunistic crowdsensing in

WWW 2017 Companion, April 3-7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3051155

social media. Our contributions can be summarized in the following:

- As a proof-of-concept, we have implemented an automatic system based on hybrid crowdsensing to demonstrate how it can help in the relevant application scenario of emergency management.
- We performed two real-world experiments and analyzed collected data in order to investigate the extent to which users are willing to collaborate and contribute information. The analysis encompasses type and time distribution of users' answers, and users' attitude to collaborate.
- Starting from a discussion of experimental results, we highlight the key factors for the success of hybrid crowdsensing as well as the main challenges to be addressed.

Even though findings reported in this study are still preliminary, we are confident that our work might lay the foundations for future developments in hybrid crowdsensing. Indeed, the proposed paradigm can be suitable for application in a number of practical scenarios including smart cities, online question answering, personalized marketing, citizen journalism, citizen science, and in all situations where a social media-based annotated dataset has to be created.

Outline. The remainder of this paper is organized as follows. Section 2 briefly surveys relevant literature. Section 3 presents our hybrid crowdsensing approach. Section 4 describes the experimental settings in which our system has been benchmarked and reports on the results. In Section 5 we critically discuss the results and implications of our experiments. Finally, in Section 6 we draw the conclusions and we highlight possible directions of future research.

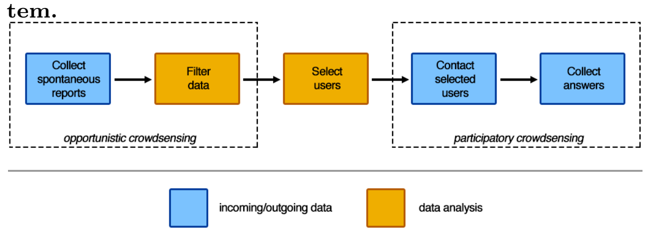
2. RELATED WORK

We survey relevant social media-based works in the fields of participatory and opportunistic sensing, and online Q&A. We conclude by discussing practically relevant social media applications.

Participatory and Opportunistic sensing. Among the first works on participatory sensing is [11]. The author describes common architectural components for participatory sensing systems and discusses technical challenges such as those of mobile device data capture, personal data stream storage, and data processing. Subsequent works such as [10, 19, 32, 22] also discuss the criticality of all participatory sensing systems regarding the need to attract a significant user base and propose possible solutions. Among these, [10, 12, 22] propose to exploit social interactions, increase user engagement with the system, and facilitate data collection and sharing in order to create and maintain a critical mass of contributors.

In addition to participatory systems, many opportunistic approaches have also been proposed in recent years. These latter type of systems builds upon data spontaneously made available by users via their daily activities, such as posting content on social networking platforms. Systems of this kind have been proposed in several scenarios such as that of emergency management [30, 28, 5, 7, 9], smart cities [29], activity and product recommendation [21], and more. In order to achieve their goals, these systems make use of powerful data analysis algorithms and techniques, with particular emphasis on data filtering. Indeed, all data gathered via opportunistic sensing is not specifically targeted to the sensing system. As such, noisy and not relevant data must be filtered out, before proceeding with the analyses.

Figure 1: General schema of a hybrid crowdsensing system.

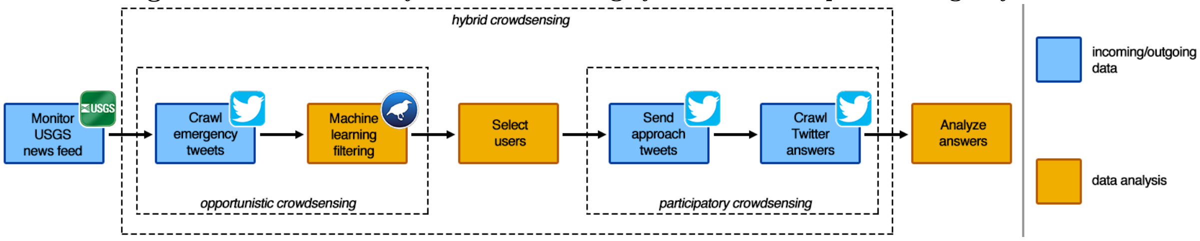


This brief survey highlighted fundamental differences between those systems based on a participatory sensing approach, with regards to those that are based on an opportunistic approach. Many participatory sensing systems are heavily focused on data collection problems and on proposing mechanisms to obtain and maintain users engagement with the system. In contrast, surveyed opportunistic systems are more focused on data analysis than data collection.

Online QA in social media. Similarly to crowdsensing, online Q&A has the goal of satisfying an information need by resorting to a crowd of experts. Works in social media Q&A typically have the goal of assessing the extent to which a social media community is willing to answer questions posed by other users. For instance, in [26] authors aim at assessing whether the Twitter community at-large may allow to obtain the same amount of information that is available within dedicated Q&A platforms like Quora and StackOverflow. Authors manually sent 1,159 questions targeting expert users and received an answer 42.3% of the times. In [27] instead, authors analyze the Twitter stream in order to measure the ratio of question tweets that receive an answer. They manually select 1,152 tweets to analyze, the 18.7% of which received an answer. This result might represent the baseline attitude of Twitter users to answering questions and might be used as a reference value in this and other works. A step forward in social media Q&A is achieved in [20], where an automatic "expert" selection process is employed for the first time. Authors perform 3 different experiments encompassing a total of 2,538 tweets and measuring an average response rate of 33.6%. To better evaluate their results, authors also compute a weak baseline by asking questions to random users. Their results show that only 3.4% of these tweets received an answer.

Our approach differs from many works in social media Q&A, in several aspects: (i) it is fully-automatic and covers an order of magnitude more tweets than previous works ($\sim 16,000$ vs. $\sim 1,100$); (ii) instead of analyzing response patterns to other users' questions and of identifying experts among community members, our goal is to propose a novel sensing paradigm and to evaluate it by implementing a fully-functional system; (iii) the questions asked in our study are sent by a centralized system rather than by other users.

Figure 2: Schema of a hybrid crowdsensing system for earthquake emergency relief.



Social media applications. Many recently developed applications are designed so as to transparently exploit available user information in order to ask targeted questions and contributions. As an example, since October 2014 Facebook has officially launched its Safety Check[1]. In the case of a major catastrophe, Facebook asks everyone within the affected area to report whether they are safe, by clicking on a button. Until recently, the Facebook Safety Check (FSC) was a manually-activated mechanism that has been used only 8 times between October 2014 and November 2015. Moreover, the FSC is specifically designed to ask a single question – whether the user is safe – in a specific scenario – that of a mass emergency. Nonetheless, it represents an interesting way of exploiting readily available data, such as user self-declared location field, to extract a set of potentially relevant users, to which the system asks a targeted question. A few months later, Google launched Local Guides[2], a Google Maps-integrated platform that helps people find, engage with, and review businesses in their local area. The app monitors users position and recognizes nearby places, prompting users to give feedback and to provide additional information.

3. HYBRID CROWDSENSING

As shown in Figure 1, hybrid crowdsensing is based on the combination of two sensing phases. In the first phase, an *opportunistic crowdsensing* module collects and filters spontaneous posts/reports from a social media with a twofold goal: (i) to figure out, as fast as possible, preliminary situational information, and (ii) to prepare a list of potential volunteers. In the second phase, a *participatory crowdsensing* module stimulates the contribution of new information by contacting selected volunteers and asking them to provide more focused and more detailed data. Based on specific application requirements, an intermediate module makes a selection of users from the list of potential contributors built in the first phase. As a result of the two sensing phases, a system implementing the proposed paradigm is able to acquire user contributions without the need of human intervention.

To better understand the potentialities of hybrid crowdsensing, we need to specify the domain-independent description given above in the context of a real life application scenario. Notably, in the last few years emergency management has been one of the favourite application domains for both opportunistic and participatory crowdsensing [15]. We argue that our approach can prove very useful in the aftermath of critical events [6], since hybrid crowdsensing can be leveraged to go beyond the event detection and collection of initial comments – achievable by an opportunistic approach alone – up to a direct contact with persons that might have experienced the event for asking them targeted questions. The latter can be achieved with a participatory approach by building upon the result of the opportunistic approach. For instance, in the case of an earthquake, detailed information about location and amount of damage could be obtained from eyewitnesses reporting from areas that are likely to be severely struck, areas scarcely covered by other sensors, or areas from which contrasting information has been received [3, 25].

4. EMERGENCY RELIEF EXPERIMENTS

In this section, we describe the experimental settings we used as a proof-of-concept and evaluation testbed for hybrid crowdsensing. The application scenario is that of emergency relief and management, with a specific focus on earthquakes, where the adoption of hybrid crowdsensing can allow for speeding up and augmenting response operations.

A hybrid crowdsensing system for earthquake emergency relief based on Twitter can be described by the schema of Figure 2. Following this schema, we implemented a fully-automatic system capable of: (i) getting a notification when an earthquake occurs, (ii) crawling Twitter based on metadata or specific keywords, (iii) filtering out noise in order to retain only relevant tweets, (iv) selecting a subset of users to be contacted in the participatory phase, (v) contacting selected users, and (vi) collecting and analyzing user replies. Henceforth, we call *approach tweet* the application-dependent tweet used to contact users.

In our experiments, the system listens to U.S. Geological Survey's (USGS) news feed, from which it obtains the notification of an earthquake occurrence as well as detailed information about its epicenter and magnitude. A new notification triggers the opportunistic crawler that, using the Twitter's Streaming API, collects as many as possible newly produced tweets that might be related to the earthquake.

The way the crawler collects such tweets defines the two different experiments we carried out. In the first one, called *geo-based* experiment, the crawler uses geographic metadata to select those tweets that were posted in the vicinity of the epicenter of the earthquake. In the second one, called *keywords-based* experiment, the crawler selects those tweets that match well-known earthquake-specific keywords [30, 5].

Geo-based experiment. Users that were in the vicinity of the epicenter of the earthquake are likely to have experienced it. As such, they represent promising targets for the participatory phase and might be asked to provide additional information on the consequences of the earthquake.

[1] `http://newsroom.fb.com/news/2014/10/introducing-safety-check/`
[2] `https://www.google.com/local/guides/`

Figure 3: Example of approach tweet used in the geo-based experiment.

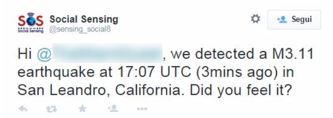


Figure 4: Examples of the 5 approach tweets used in the keywords-based experiment.

To reach these users, we can exploit tweet *geolocation* – a feature that allows users to disclose the GPS coordinates of their current position – to select users that posted in the aftermath of an earthquake while lying within an appropriate range from the epicenter ($\sim$ 15km/9.3miles[3]).

As shown in Figure 3, the approach tweet to these users asks whether they actually felt the shake. Notably, such a simple yet friendly approach is also used in a well established system operated by the USGS to infer the severity of earthquakes, where registered users are sent emails asking them to fill a Web-enabled damage report. In a practical use case, our hybrid crowdsensing system could automate the existing USGS's tool by asking the data for the report directly to all the Twitter users in the vicinity of the epicenter, without the constraint of relying only on already registered users. Another advantage of our approach could be improved responsiveness due to the lower time required to obtain answers via Twitter compared to that of Web surveys or emails.

Keywords-based experiment. This experiment is motivated by recent statistics reporting that only up to 4% of tweets are geolocated [3]. Thus, to address the scarcity of geolocated Twitter emergency reports, in the opportunistic phase we can seek tweets posted in the aftermath of an earthquake and containing specific keywords, such as "earthquake" or "quake". The rationale is that users posting those tweets might have experienced the earthquake and thus represent potential targets for our participatory phase [30, 5]. Unfortunately, because of the lack of information about an user's position, we can not distinguish between users that actually felt the earthquake and the ones that incidentally used a selection keyword with no relation to the event. This suggested us to use different approach tweets that, instead of asking whether a user felt the shake, pose more dubitative questions. Figure 4 shows the different types of approach tweets we used in this experiment.

Noise filtering. For both experiments, the opportunistic phase must be refined by a filtering step in order to exclude

[3]The precise range may vary slightly depending on the magnitude and the depth of the hypocenter.

Class	TP Rate	FP Rate	Precision	Recall	F-Measure
related	0.854	0.064	0.893	0.854	0.873
not related	0.936	0.146	0.911	0.936	0.924
weighted average	0.905	0.115	0.904	0.905	0.904

Table 1: Noise filtering: performance of the machine learning classifier.

the tweets not related to the earthquake that just occurred. Following the successful approach of previous works [30, 5], we used a machine learning classifier to infer whether a tweet is relevant or not. The classifier has been implemented in Weka using the decision tree J48 and trained with manually labeled tweets[4].

Selecting users. In principle, before initiating the participatory phase we should select potential contributors from the list of users that posted at least one relevant tweet among those collected during the opportunistic phase. For instance, this selection task can be performed through advanced witness detection techniques proposed in literature [25]. However, given the difficulty in repeating the experiments (i.e., the need of a major earthquake occurrence), to have better statistics we preferred to contact all available users that posted at least one relevant tweet. Then, in the participatory phase, the system automatically contacts users by replying to their original tweets with the approach tweets described above and, by means of a second Twitter streaming crawler, collects possible users' answers to the approach tweets.

Experimental settings. The emergency relief experiments took place between February and May 2015. It should be noted that the real-time nature of these experiments – they can be carried out only in the presence of a real disaster – requires long-lived crowdsensing campaigns and prevents us from planning the real amount of collectable data. Nevertheless, during that time span, 931 earthquakes having magnitude ≥ 2.5 occurred worldwide, according to the USGS. In the aftermath of those earthquakes we contacted $\sim 16,000$ users. More than 3,600 of those users answered to our questions, overall generating $\sim 5,800$ replies[5]. In all the experiments, no user has been contacted more than once.

4.1 Noise filtering results

Table 1 shows the performance of the Weka classifier used to filter out noise and retain only relevant tweets. The classifier was trained on a set of 5,469 manually labeled tweets with a 10-fold cross validation, taking in features such as the length and publication timestamp of the tweet, and the presence on its content of capital letters, specific punctuation characters and mentions. Overall, the classifier is rather accurate, as demonstrated by a solid F-Measure of over 0.9.

4.2 Willingness to answer results

Similarly to what is done in crowdsourcing systems such as Amazon Mechanical Turk and CrowdFlower, the typical application scenario of hybrid crowdsensing involves asking users for their help. The difference is that, instead of having users that deliberately join the crowdsourcing platform,

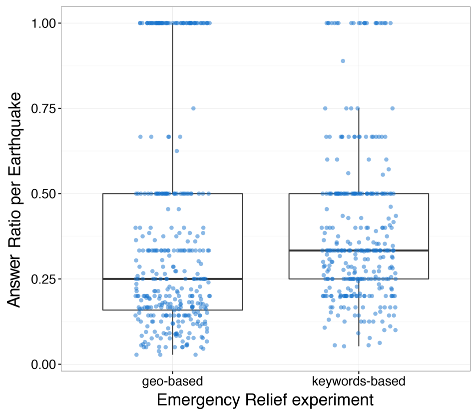
[4]Subsequent Section 4.1 provides more detailed information.

[5]Since a few users answered our questions with more than one reply tweet: num. replies > num. answers.

Description	# questions	# answers	answer ratio
experiments			
Geo-based	6,636	1,259	19.0%
Keywords-based	9,332	2,459	26.4%
baselines			
Questions to random users [20]	476	16	3.4%
Spam tweets [8]	117,192	12,943	11.0%
Questions from Twitter users [27]	1,152	215	18.7%

Table 2: Willingness to answer: results of emergency relief experiments vs. baselines.

Figure 5: Emergency relief experiments: boxplot and scatterplot distributions of answer ratios per earthquake.



with hybrid crowdsensing we directly contact them. Thus, the usefulness of the hybrid crowdsensing paradigm heavily relies on users' willingness to collaborate and, ultimately, on the number of answers to the approach tweets.

To understand this key point, we need statistical baselines of the behaviors of social media users in reaction to messages asking them to perform some simple task. In addition to the baselines derived from recent research in Q&A, we obtained a new baseline by measuring the users' reaction to messages sent by a set of advanced Twitter spambots, thoroughly studied and described in [8]. Such spambots perform their malicious activities by mentioning random Twitter users in their automatically created tweets and by inviting them to buy a paid app from the Apple Store. Intuitively, messages shared by such spambots are of little interest to Twitter users and hence we expect a rather low answer ratio to the tweets sent by spambots. Nonetheless, we are interested in verifying if users are more willing to answer to our approach tweets than they are at answering to spambots.

The willingness of individuals to reply to our requests can be measured by the percentage of approach tweets that received an answer. Table 2 shows the *answer ratios* obtained for the geo-based and keywords-based experiments (19.0% and 26.4%, respectively) as well as those measured for 3 baselines (3.4%, 11.0%, and 18.7%). As a result of a chi-squared test (χ^2), all the differences between answer ratios of our experiments and those of the baselines proved to be statistically significant ($p < 0.01$).

Figure 5 shows a fine-grained analysis of answer ratios for the two emergency relief experiments, separately. Each dot in the plot is related to the answer ratio measured for a single earthquake, with all the dots in the plot summing up to the 931 earthquakes covered by our study. Results confirm those of Table 2, with the median answer ratio for

Figure 6: Keywords-based experiment: boxplot and scatterplot distributions of answer ratios per question.

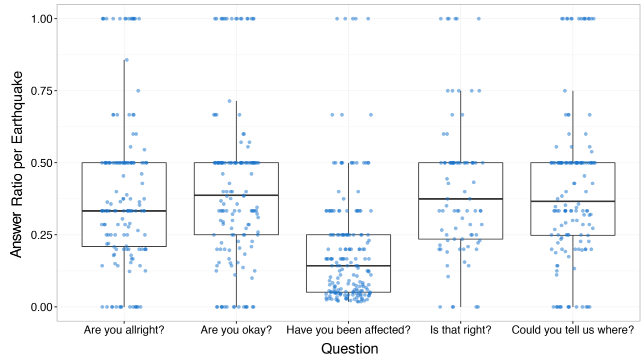
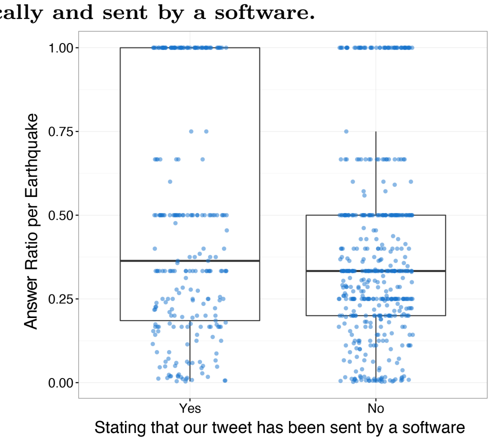


Figure 7: Keywords-based experiment. Boxplot and scatterplot distributions of answer ratios based on whether we specified that our tweet was created automatically and sent by a software.



the keywords-based experiment moderately higher than the geo-based one. As discussed in the next section, this 7.4% difference in the reply ratio ($p < 0.01$), is both substantial and highly significant.

In order to understand how the type of question might influence the willingness to answer, in Figure 6 we show the answer ratios obtained with the five approach tweets used in the keywords-based experiment (see Figure 4). The distribution of answer ratios is quite uniform, with the only exception of the question *"Have you been affected?"*. The differences between the answer ratios measured for the *"Have you been affected?"* question with regards to all other questions, are statistically significant, with all $p < 0.01$.

Finally, we investigated whether clearly stating that the approach tweet was automatically generated could change user answering behaviors. Figure 7 shows the boxplot and scatterplot distributions of answer ratios in the two cases. The distributions appear very similar, with the boxes completely overlapping and comparable median values. Moreover, a χ^2 test of statistical significance of the difference between the two answer ratios resulted in $p = 0.065$, thus failing to reject the null hypothesis at the 5% significance level. Both Figure 7 and the significance test seem to support the claim that there is no statistically significant dif-

Figure 8: Probability of receiving an answer to the approach tweets as a function of the delay with which the system contacts users.

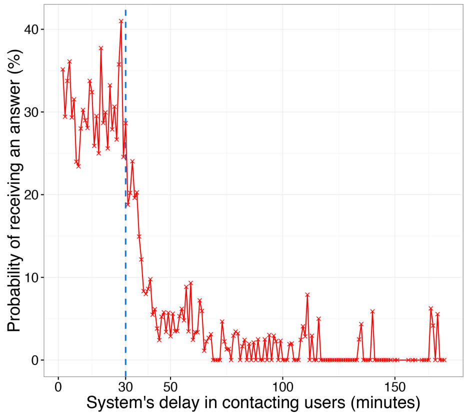
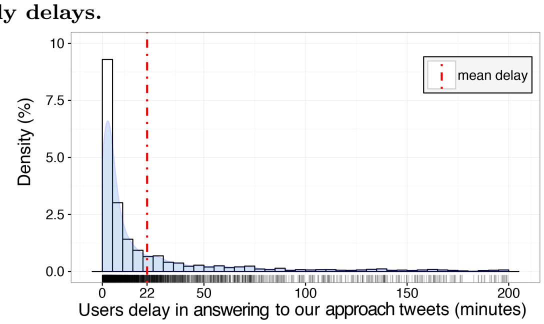


Figure 9: Probability density histogram and rug plot of reply delays.



ference in answer ratio between stating or omitting that our messages are automatically generated and sent.

4.3 Time distribution of answers

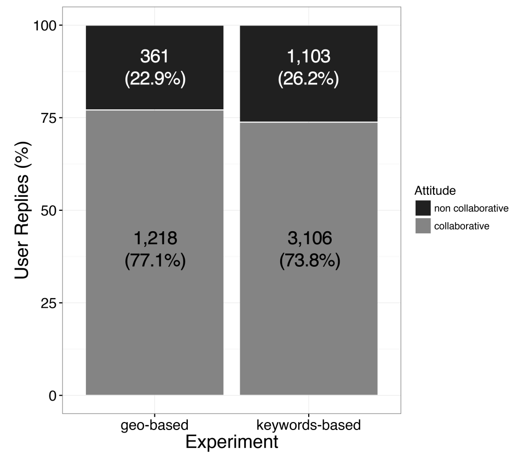
We observed a significant decrease in the number of users' answers in relation to the delay with which our system sends the approach tweet – this delay being measured from the time users posted their messages. By comparing the numbers of approach tweets that received an answer to those that did not, as a function of the delay in sending the approach tweet, we obtain the plot of Figure 8. The plot confirms the intuition that the later the system contacts a user, the lower is the chance of receiving an answer. Interestingly, the probability of receiving an answer remains almost uniform, in the region of 30%, within the first 30 minutes from a user's tweet. For approach tweets sent with a delay between 30 and 60 minutes, and for those sent more than 1 hour later, the probability of receiving an answer rapidly drops at around 5% and 1%, respectively.

Another interesting experimental finding is the distribution of delays with which users answer to our approach tweets. Figure 9 depicts the probability density histogram of answer delays, which shows that the probability of receiving an answer decreases exponentially with the time. The mean value of the delay is ~ 22 minutes (represented by the dot-dashed vertical red line). For instance, in the aftermath of an emergency it is important to obtain information about the unfolding situation as fast as possible.

		A1		
		collaborative	non collaborative	not available
A2	collaborative	**4,063**	327	7
	non collaborative	203	**1,191**	7
	not available	6	10	**34**

Table 3: Results of the human annotation task. Agreement between annotators is marked with bold font.

Figure 10: Analysis of cooperativeness in replies to our tweets.



4.4 User cooperativeness

More interesting than a mere analysis of the number and timing of the answers obtained from contacted users, it is an analysis of the *content* of those answers and replies. Notably, our analysis is more detailed than those typically carried out in previous social media Q&A studies [24, 27, 26, 20, 18].

In order to assess whether users that answered to our approach tweets were actually willing to collaborate, we need to figure out to which extent they provided useful answers to our questions. Then, we asked two human annotators[6] to classify as "collaborative" or "non collaborative" all the 5,848 replies received in our two experiments. We consider as "collaborative" those replies that aim at answering our questions (see Figures 3 and 4) and provide help. Conversely, "non collaborative" are replies that do not convey any useful information, such as those where users make jokes about our approach tweets. For this annotation task we also included the "not available" class in order to mark tweets that got deleted by their authors after we collected the replies. Results of the annotation task are reported in Table 3 with annotators labeled `A1` and `A2`. As shown in table, the two annotators agreed in their classification in 90.4% of all the replies, which results in an excellent Cohen's kappa [13] inter-annotator agreement $\kappa = 0.7532$.

In order to discriminate those 560 replies (9.6%) where `A1` and `A2` had contrasting judgements, we let a super-annotator solve the discrepancies and decide the final label. Moreover, we discarded from our analysis 64 replies (1.1%) that were labeled as "not available" by either annotators.

Final results of this analysis are shown in Figure 10. For both the geo-based and the keywords-based emergency relief experiments the majority of replies is collaborative: 77.1% and 73.8%, respectively.

[6] Our annotators are post graduate students in Computer Science with yearly experience as social media users.

	Original tweet	Approach tweet	Answer tweet
(a)	Earthquake oh my gosh steel feeling	Hi <*anonimized*>, @socialsensing has seen you may have been involved in an earthquake. Are you alright?	@cnrsocial9 @socialsensing the earthquake has not gone yet. Situation is worst here
(b)	Earthquake in Delhi...	Hi <*anonimized*>, @socialsensing has seen you may have been involved in an earthquake. Are you alright?	@cnrsocial6 @socialsensing yes we are all fine. Thank you so much
(c)	Earthquake! We just had a small one here in Santa Barbara...	Hi <*anonimized*>, @socialsensing has seen you may have felt an earthquake. Could you tell us if you've been affected?	@cnrsocial13 @socialsensing No damage here, near downtown and the Mission in Santa Barbara
(d)	We all r fine after the earthquake.	Hi <*anonimized*>, @socialsensing has seen you may have been involved in an earthquake. Are you alright?	@cnrsocial10 @socialsensing we all r fine here. thanks for your concern.
(e)	wait. 1.01 no wonder I didn't feel the earthquake	Hi <*anonimized*>, @socialsensing has seen you may have been involved in an earthquake, could you tell us where?	@cnrsocial9 @socialsensing I'm around the Los Angeles area, I don't want to tell the precise location tho
(f)	Deadly earthquake Nepal - At least 1.989 people killed - very strong NEW EARTHQUAKE East of Kathmandu -	Hi <*anonimized*> this is an auto-response. We have noticed you may have been involved in an earthquake, could you tell us where are you?	@cnrsocial great attempt to get exps like that. Cooperating might not be a bad idea
(g)	earthquake.	Hi <*anonimized*>, @socialsensing has seen you may have been involved in an earthquake. Is everything okay?	@cnrsocial12 yes! Everything is okay now :)
(h)	Whoa, bit an #earthquake tremor south of #dallas.	Hi <*anonimized*>, @socialsensing has seen you may have been involved in an earthquake, could you tell us where?	@cnrsocial3 @socialsensing Mansfield, TX. Very minor, desk shook.
(i)	Oh look. Another ring of fire 7+ quake. Fffffff	Hi <*anonimized*>, @socialsensing has seen you may have been involved in an earthquake. Are you alright?	@cnrsocial11 @socialsensing I wasn't in an earthquake, I was merely commenting on recent quakes. Your algorithm needs a little tweaking :)
(j)	Now playing - Nightstep - Earthquake	Hi <*anonimized*>, @socialsensing has seen you may have been involved in an earthquake, could you tell us where are you?	@cnrsocial12 @socialsensing no that's the name of a song

Table 4: Examples of conversations with Twitter users. As shown, our questions obtained a broad set of different answers.

5. DISCUSSION

In this section we carry out a thorough discussion on the key results of our experiments. We also highlight challenges, as well as success and risk factors of hybrid crowdsensing systems.

Emotional involvement is a catalyst for contribution. In our emergency relief experiments we measured an answer ratio ranging from 19.0% (geo-based experiment) to 26.4% (keywords-based experiment). These results are promising and support the feasibility of our proposed approach, also considering that our proof-of-concept system is capable of collecting such answers in a fully automated fashion. This moderate and significant difference in the answer ratios between the two experiments can be explained by considering the different types of questions asked to the users. The approach tweet question in the geo-based experiment is formal and based on objective information, since it reports the magnitude, time, and precise location of the epicenter. Hence, it is worded so that it might feel lacking emotional concern towards affected users. Instead, the more generic and sympathetic questions of the keywords-based experiment could better motivate users to answer. In turn, we believe that this result highlights a fundamental characteristic of the so-called "social sensors", namely they are more willing to provide contributions for topics in which they are emotionally involved. A strong emotional component in user replies is evident from the answers in Table 4 (b), (d), and (g). This finding is also supported by a comparison between the answer ratio measured in our experiments and those reported in previous related works, especially those in the Q&A field. For instance, all answer ratios to generic questions reported in [27, 8] are lower than ours. Instead, works that achieved higher answer ratios did so via a manual contacting process [26, 20], or by selecting only a subset of users that were more likely to answer [18, 20]. The finding that emotional involvement is a key contribution factor for users also implies that deploying a hybrid crowdsensing system in different practical scenarios could possibly result in a different number of received contributions.

The quest for the perfect question. The analysis of the answer ratios to the 5 different questions asked within the keywords-based experiment showed no significant differences, with the only notable exception of the question "*Have you been affected?*", for which we measured a significantly lower ratio of answers. This result highlights the importance of "question design" – that is, the task of designing the questions – and the accounts that ask those questions – in such a way to maximize the number and quality of received answers. To this end, we envision the possibility to leverage previous experiences in fields such as social engineering [14], crowdsourcing [23], and even psychology [31].

Responsiveness of hybrid crowdsensing. Among the most important variables to consider in crowdsourced systems, and especially those related to emergency management, is the timing. Results about the probability of receiving an answer to our questions as a function of the delay in contacting users, showed that within the first 30 minutes of a user's tweet there is a $\sim 30\%$ chance of obtaining an answer. Additional results showed that the answer is likely to be provided within the first few minutes, with the majority of answers arriving within the first 5-10 minutes from our approach tweets. These findings support our initial claim that a system based on hybrid crowdsensing can retain the timeliness and large user base typical of opportunistic systems, while being able to collect detailed and targeted information that is typical of participatory systems. In addition, these results also support the adoption of hybrid crowdsensing in real-time applications and in practical scenarios imposing stringent time requirements.

Collaborative social sensors. The success of a system based on hybrid crowdsensing necessarily depends on managing not to bother users. In contrast with crowdsourcing systems, where users register and willingly decide to contribute to specific tasks, our paradigm asks for contributions targeting almost "unaware" users. As such, we might expect that a fraction of all the replies we receive are unfavorable and unhelpful. However, a manual analysis of all the 5,848 replies received in our 2 experiments showed that on average 74.7% of all replies were collaborative. This means that the vast majority of replies were sent with the intent of providing useful information. In addition, also those users that did not provide useful information, did so in an overall positive manner, such as in the example of Table 4 (f). Unfortunately, much of the previous literature in this field focused on the answer ratio and on the time distribution of answers, rather than on their content. For this reason it is difficult to cross check these results. One notable exception is [27] that examined if the replies to a random set of questions asked by Twitter users were relevant or not. Authors found that 84% of the responses were relevant. Our results are in line with those of [27], also considering that the questions studied in [27] were posed by real Twitter users to the community. Our questions instead targeted unaware users, that might have not been willing to provide help.

Trade-offs and limitations. The study of user replies also highlighted a crucial trade-off of hybrid crowdsensing. On the one hand, there is the need to acquire as much useful information as possible, as fast as possible. On the other hand, a system based on hybrid crowdsensing has to cope with user reactions and privacy issues.

User reactions and their perception of the contacting system are important, since spamming and annoying users might imply the failure of current and future sensing campaigns. To this regard, during the 4 months of our experiments, we deliberately constrained the rate at which we sent our approach tweets. We also decided to contact each user at most once, in order to avoid spamming. We believe these, and other, considerations to be important towards receiving positive and constructive contributions. Our results to this regard are comforting and users showed to perceive our system in a positive and collaborative way, as shown in Table 4.

In our experiments, contacted users did not seem to be particularly concerned about their privacy and were keen to share detailed information with us – see for example Table 4 (h), (a), and (c). Nonetheless a minority of users, although willing to help, were reluctant to share precise information about themselves. This was particularly evident when asking users about their precise location, as shown in Table 4 (e).

6. CONCLUDING REMARKS

The proposed hybrid crowdsensing paradigm, lying at the intersection of traditional crowdsensing and online Q&A, allows to draw upon a large user base while still retaining the chance to obtain targeted, high quality, and detailed information. Furthermore, it enables a fast data collection without requiring any human intervention.

Our findings demonstrate that automatically generated questions are answered up to 30% of the times, with the vast majority of answers given within the first 5-10 minutes, $\sim 75\%$ being collaborative. Thus, we can argue that hybrid crowdsensing allows to obtain rich and detailed information, which are typical of participatory systems, while still having access to a large user base of contributors, which is typical of opportunistic systems.

Regarding future improvements of hybrid crowdsensing systems, rather than sending one single question to a large number of users, a possibly more efficient approach would require to contact a subset of users that are known (or are predicted) to be willing to respond and cooperate, and ask them several questions. This task resembles that of "expert finding" in online Q&A communities and, in the near future, previously proposed techniques in Q&A might be benchmarked in the hybrid crowdsensing context. Finally, as suggested by the mistakes shown in Table 4 (i) and (j), future versions of our system will need to improve in filtering data and selecting users to contact.

7. ACKNOWLEDGEMENTS

This research is supported in part by the EU H2020 Program under the scheme `INFRAIA-1-2014-2015: Research Infrastructures` grant agreement #654024 *SoBigData: Social Mining & Big Data Ecosystem*, by the `PRA 2016` project *Analysis of Sensory Data: from Traditional Sensors to Social Sensors* funded by the University of Pisa, and by the MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca) and Regione Toscana (Tuscany, Italy) funding the *SmartNews: Social sensing for Breaking News* project: `PAR-FAS 2007-2013`.

Regione Toscana | FAS Fondo Aree Sottoutilizzate 2007-2013 | REPUBBLICA ITALIANA

8. REFERENCES

[1] C. C. Aggarwal and T. Abdelzaher. Social sensing. In *Managing and Mining Sensor Data*. Springer, 2013.

[2] M. Avvenuti, M. G. Cimino, S. Cresci, A. Marchetti, and M. Tesconi. A framework for detecting unfolding emergencies using humans as sensors. *SpringerPlus*, 5(1):1–23, 2016.

[3] M. Avvenuti, S. Cresci, F. Del Vigna, and M. Tesconi. Impromptu crisis mapping to prioritize emergency response. *IEEE Computer*, 49(5):28–37, 2016.

[4] M. Avvenuti, S. Cresci, M. N. La Polla, A. Marchetti, and M. Tesconi. Earthquake emergency management by social sensing. In *PERCOM'14 Workshops*.

[5] M. Avvenuti, S. Cresci, A. Marchetti, C. Meletti, and M. Tesconi. EARS (Earthquake Alert and Report System): A real time decision support system for earthquake crisis management. In *KDD'14*.

[6] M. Avvenuti, S. Cresci, A. Marchetti, C. Meletti, and M. Tesconi. Predictability or Early Warning: Using Social Media in Modern Emergency Response. *IEEE Internet Computing*, 20(6):4–6, 2016.

[7] S. Cresci, A. Cimino, F. Dell'Orletta, and M. Tesconi. Crisis mapping during natural disasters via text analysis of social media messages. In *WISE'15*.

[8] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi. DNA-inspired online behavioral modeling and its application to spambot detection. *IEEE Intelligent Systems*, 31(5):58–64, 2016.

[9] S. Cresci, M. Tesconi, A. Cimino, and F. Dell'Orletta. A linguistically-driven approach to cross-event damage assessment of natural disasters from social media messages. In *WWW'15 companion*.

[10] A. L. Dallora Moraes, F. Fonseca, M. G. P. Esteves, D. Schneider, and J. M. de Souza. A meta-model for crowdsourcing platforms in data collection and participatory sensing. In *CSCWD'14*.

[11] D. Estrin. Participatory sensing: applications and architecture [internet predictions]. *IEEE Internet Computing*, 14(1):12–42, 2010.

[12] J. A. Fails, K. G. Herbert, E. Hill, C. Loeschorn, S. Kordecki, D. Dymko, A. DeStefano, and Z. Christian. GeoTagger: a collaborative and participatory environmental inquiry system. In *CSCW'14*.

[13] K. L. Gwet. *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC, 2014.

[14] C. Hadnagy. *Social engineering: The art of human hacking*. John Wiley & Sons, 2010.

[15] M. Imran, C. Castillo, F. Diaz, and S. Vieweg. Processing social media messages in mass emergency: a survey. *ACM Computing Surveys*, 47(4):67, 2015.

[16] A. Kapadia, D. Kotz, and N. Triandopoulos. Opportunistic sensing: Security challenges for the new paradigm. In *COMSNETS'09*.

[17] N. D. Lane, S. B. Eisenman, M. Musolesi, E. Miluzzo, and A. T. Campbell. Urban sensing systems: opportunistic or participatory? In *HotMobile'08*.

[18] Y. Liu, T. Alexandrova, and T. Nakajima. Using stranger as sensors: temporal and geo-sensitive question answering via social media. In *WWW'13*.

[19] T. Ludwig, T. Siebigteroth, and V. Pipek. Crowdmonitor: Monitoring physical and digital activities of citizens during emergencies. In *SocInfo'14*.

[20] J. Mahmud, M. X. Zhou, N. Megiddo, J. Nichols, and C. Drews. Recommending targeted strangers from whom to solicit information on social media. In *IUI'13*.

[21] J. Mayer and Q. Jones. Encount'r: Exploring Passive Context-Awareness for Opportunistic Social Matching. In *CSCW'16 companion*.

[22] E. Medina, D. Lopez, D. Royo, R. Meseguer, and S. F. Ochoa. CoSP: A Collaborative Sensing Platform for Mobile Applications. In *CSCWD'15*.

[23] P. Michelucci. *Handbook of human computation*. Springer, 2013.

[24] M. R. Morris, J. Teevan, and K. Panovich. What do people ask their social networks, and why?: a survey study of status message Q&A behavior. In *CHI'10*.

[25] F. Morstatter, N. Lubold, H. Pon-Barry, J. Pfeffer, and H. Liu. Finding eyewitness tweets during crises. In *ACL'14*.

[26] J. Nichols and J.-H. Kang. Asking questions of targeted strangers on social networks. In *CSCW'12*.

[27] S. A. Paul, L. Hong, and E. H. Chi. Is twitter a good place for asking questions? a characterization study. In *ICWSM'11*.

[28] B. Robinson, R. Power, and M. Cameron. A sensitive twitter earthquake detector. In *WWW'13 companion*. ACM.

[29] H. Roitman, J. Mamou, S. Mehta, A. Satt, and L. Subramaniam. Harnessing the crowds for smart city sensing. In *CrowdSens'12*.

[30] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *WWW'10*.

[31] N. Schwarz, B. Knäuper, D. Oyserman, and C. Stich. The psychology of asking questions. *International handbook of survey methodology*, pages 18–34, 2008.

[32] A. Tomasic, J. Zimmerman, A. Steinfeld, and Y. Huang. Motivating contribution in a participatory sensing system via quid-pro-quo. In *CSCW'14*.

[33] J. Zhang, M. S. Ackerman, and L. Adamic. Expertise networks in online communities: structure and algorithms. In *WWW'07*.