Veracity and Velocity of Social Media Content during Breaking News: Analysis of November 2015 Paris Shootings

Stefanie Wiegand and Stuart E. Middleton
University of Southampton IT Innovation Centre
Gamma House, Enterprise Road,
Southampton SO16 7NS, UK
tel: +44 23 8059 8866
email: {sw,sem}@it-innovation.soton.ac.uk

ABSTRACT
Social media sources are becoming increasingly important in journalism. Under breaking news deadlines semi-automated support for identification and verification of content is critical. We describe a large scale content-level analysis of over 6 million Twitter, You Tube and Instagram records covering the first 6 hours of the November 2015 Paris shootings. We ground our analysis by tracing how 5 ground truth images used in actual news reports went viral. We look at velocity of newsworthy content and its veracity with regards trusted source attribution. We also examine temporal segmentation combined with statistical frequency counters to identify likely eyewitness content for input to real-time breaking content feeds. Our results suggest attribution to trusted sources might be a good indicator of content veracity, and that temporal segmentation coupled with frequency statistical metrics could be used to highlight in real-time eyewitness content if applied with some additional text filters.

Keywords
Velocity; Veracity; Trust; Credibility; Natural Language Processing; Semantics; Social Media; User Generated Content; News

1. INTRODUCTION
Social media sources are becoming increasingly important and pervasive within the field of journalism. Before user generated content (UGC) can be used it must be identified, verified and integrated into the final news story. In breaking news the deadlines are measured in minutes not days, so tools which can automate parts of this process are very important when journalists are confronted with millions of possible social media content items to filter and analyse. Competition is fierce between journalists and kudos goes to the one who publishes breaking news first; shortcutting verification steps however risks publishing rumour as truth and can ruin a journalist's professional reputation. It is important that journalists find and verify user generated content quickly and correctly.

Of particular interest to breaking news stories are images and videos from eyewitnesses at the scene of an event. These are often uploaded to media sharing sites and/or social media sites as the news event unfolds. Journalists find these images and videos by monitoring known sources (e.g. real-time feeds based on curated sources in Twitter lists), trending content (e.g. searches based on Twitter trending hashtags) and active keyword searches (e.g. Twitter search). Once an eyewitness image or video is found a journalist will typically [1] look for the original posting, contact the author of this post and then ask some questions to verify that they are indeed the true author. Lastly they will ask for permission to publish the content in their news story.

Dashboard applications such as TweetDeck[1] and Storyful[2] make it easier for users to manage real-time streams of social media content with the hope of finding eyewitness and newsworthy content soon after it is posted. These tools allow management of multiple keyword filtered streams. Image search tools such as TinEye[3] and Google Reverse Image Search[4] are used by journalists to find duplicates, such as other posts of same image, and near duplicates, such as posts before or after potential Photoshop manipulations, to help find fake posts. If the original image is located its metadata can be used to extract facts, for example the make and model of image recording device, which can be confirmed by the author when they are contacted.

Approaches to faking [2] include digital manipulation, recycling old content as breaking news and misrepresentation of genuine content. For example TV banners can be added to make a video look like it came from a trusted source, old war footage recycled in more recent conflicts or images from innocent accidents presented in ways that support claims of foul play. A journalist will ultimately use human judgement to decide if content items are genuine or not, however tools can help a lot in this process by providing relevant contextual evidence to base decisions upon.

Understanding the dynamics of how newsworthy content goes viral is important to developing better tools to support the identification and verification processes. We present in this paper our analysis of the first 6 hours of the November 2015 Paris shootings, looking at content crawled from Twitter, You Tube and Instagram. Our qualitative analysis is anchored to 5 ground truth social media images that appeared in broadcast news stories during the event, including 3 genuine images and 2 fake images that were debunked shortly after being broadcast. To look at the velocity of newsworthy content we temporally segment our dataset and show metrics for mentions of ground truth images over time. We report

WWW 2016 Companion, April 11-15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
DOI: http://dx.doi.org/10.1145/2872518.2890095

[1] https://tweetdeck.twitter.com
[2] https://storyful.com
[3] https://www.tineye.com
[4] https://images.google.com

metrics for both the original image and duplicates found using TinEye. With regards to veracity we look at which percentage of these mentions were from either trusted sources directly, or attributed to trusted sources indirectly. We lastly look at supporting real-time eyewitness content identification by examining how temporal segmentation, combined with statistical frequency counters, can be used to find real-time lists of original breaking content which are likely to contain eyewitness images and videos. This latter approach has the potential to dramatically reduce the volume of content journalists need to monitor, allowing them more time to get the verification work finished.

The novelty of this work derives from our large scale analysis of the first hours of a real breaking news story, as opposed to time periods well after the story breaks which is much more common in the published literature. Previously published large scale work involving breaking news typically involves traffic analysis and user mention graphs, and nothing at the deeper content-level. Published work at a deeper content-level focuses on benchmark datasets, small in size (i.e. only thousands of content items) and usually manually extracted and labelled. We provide a large scale content-level analysis of 38 GB of content, fully covering the first 6 hours of the Paris shootings event, which we hope practitioners and researchers can use in the future to help guide news analysis tool development.

We outline in section 2 related work and our analysis approach in section 3. The experiment setup and results are described in section 0, with discussion and conclusions in sections 5 and 6.

2. RELATED WORK

Published social media analytics during news events is mostly focussed on Twitter, using data from traffic analysis and sentiment analysis techniques to look into specific case studies. An example of analysis on a breaking news story is [6], where keywords are used to look at tweet sentiment (e.g. certain, uncertain) in reports of the death of Osama Bin Laden in 2010 both before, during and after the event. This analysis is small-scale (i.e. about 900 tweets) using a manually labelled dataset. A larger-scale analysis [7] looked at 4 million tweets from 5,000 sources crawled during the 2012 US election. This work analysed temporal traffic metrics (e.g. tweets per minute) during key political events such as Barack Obama's victory tweet. Another large scale analysis [8] looked at the 2011 UK summer riots, analysing 2.6 million tweets. This work showed that journalists and mainstream media posted the majority of content with a long tail of 'silent majority' readers, with an in-depth analysis of posts from two 'at the scene' journalists providing a qualitative insight into how eyewitness media reports went viral.

Some approaches use visualizations to help users trace back content to the original post. An example is [9] where clustered tweet propagation from a target tweet is displayed on a timeline. This system uses the Twitter Search API, and is limited to data crawled within the 7 day window in the same way our work is. Another work [10] has examined overlaying social network interconnections to temporal graphs of rumour retweets, revealing active users in both graphs during propagation periods as the rumours goes viral. These works lack deep content-level analysis, such as the extraction of attributed sources that we show.

3. METHOD

We used a number of previously developed analytics tools to crawl and process the Paris shootings datasets, as described in Figure 1. This paper only reports the results of our analysis work, and does not contain details of the technical approaches used; interested readers can find such details in [3], [4] and [5].

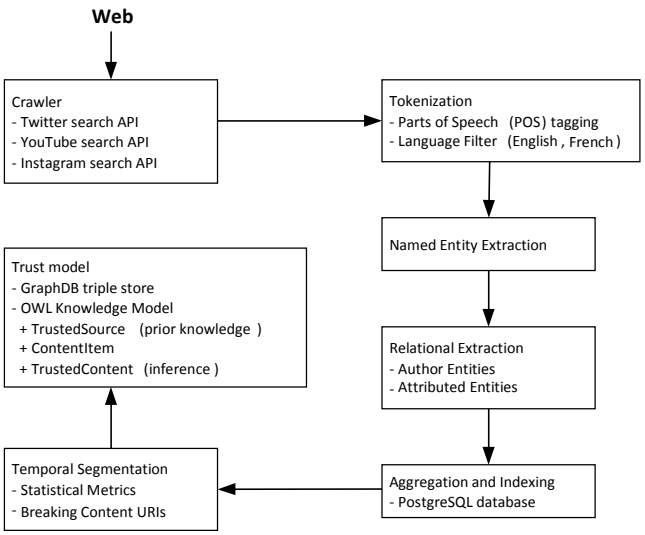

Figure 1. Analytics Technical Workflow

We crawled our dataset using the Twitter, YouTube and Instagram search APIs. We used our own crawler software [4] [5] with Twitter hashtag filters of '#Paris' and '#ParisShootings' and YouTube/Instagram location filters for Paris. Since we crawled within 7 days of the Paris shootings this allowing us to execute full historical searches, gradually paging the results back in time until the target start date. We extracted 6 hours (i.e. 38 GB serialised json) of historical content this way, including full coverage of the event start time (i.e. a period from 13-11-2015 20:20:00 UTC to 14-11-2015 02:00:00 UTC). The volume of social media content we obtained this way is much larger than is available using the Twitter Streaming API, which only runs on a small sample of the firehose, and we were able to approach levels available to services with full Twitter firehose access. The aggregated and indexed data contains 5.9M content items from 2.4M authors, 1.2M of which were attributed to 40k named entities. 418k unique URLs were shared in 4M content items.

Our analysis software processed the JSON metadata for each content item, extracting the timestamp, author, media and textual components. The text went through a natural language processing pipeline [3], involving named entity extraction and relational extraction, to extract mentions of attributed entities like 'BBC News'. Each content item was then stored in a PostgreSQL database and cross-indexed to each extracted entity (i.e. author, attributed entity, media links). This then allowed SQL queries to be executed to temporally segment the dataset (e.g. 5 minutes segments) and return ranked lists of trending authors, entities and media links for each temporal segment. We finally imported temporal segments of the data into a knowledge-based model we have created. This knowledge-model associates authors and attributed entities to a-priori declared trusted sources, allowing different levels of trusted content items to be inferred. The import of each 1 hour segment took between 15 minutes and 3 hours (not optimised), depending on the data. This figure can be significantly improved by using a machine with more memory.

4. EXPERIMENT

4.1 Experiment setup

For our ground truth we compiled a list of (un)trusted sources, i.e. sources we defined in to be either trusted or untrusted. News organisations such as BBC Breaking have in the range of 30 specialised lists of sources[5], each with about 200 names. For our qualitative evaluation we created a list of 49 trusted and 18 untrusted sources, using sources which appeared with a high frequency in our dataset to reduce the manual effort in creating the list. Our trusted sources focus on large news channels such as BBC or CNN. Our untrusted sources are smaller news agencies or individual journalists with a history of spreading false rumours. We import these lists into our knowledge base along with the actual content. Trust related information is stored as a separate RDF graph to represent a viewpoint, since different journalists would have different trusted source lists for different purposes (e.g. one list for news related to American politics, another one for international sports or military operations).

4.2 Experiment method

For each test case we imported relevant content items from a PostgreSQL indexed content database for the crawled news story into a GraphDB triple store. Then we used our trust model application to run queries on the database and/or triple store to generate results.

For each of the 5 target images as shown in table 1 we created an expanded set of URIs consisting of the original post and duplicates found using a TinEye reverse image search. Each expanded list of URIs was then used to filter content in our queries to only those posts embedding or linking to the target images. All of these pictures appear in ground truth news articles with and without attribution to the original author.

Table 1. Statistical information about the target images (each originally embedded in a tweet), their authors and whether it was crawled as part of our dataset or just mentioned by other content.

Target Image ID	P1	P2	P3	D1	D2
author is a journalist / news org	n	y	n	y	y
number of followers of author	335	1.4k	218	2.8k	151k
content likes	11	408	35	17k	29k
content retweets	83	3.3k	194	22k	30k
originally crawled	n	y	n	y	y
total # of tweets in 60 minute window	483918	162111	811079	1501000	1837173
total # of unique mentioned URLs in 60 minute window	785	4331	535	7907	13252

In our first experiment we ran queries on 10 minute temporal segments starting from the first mention of each target image in our dataset. This data was imported into our trust model allowing trusted authors, and attributions to trusted sources to be analysed. The aim of this experiment was to examine the first hour of content mentioning each target image, breaking it down into total mentions and mentions from, or attributed to, (un)trusted sources. This is relevant to journalists trying to identify verified content soon after it is published, which might have contextual relevance to an event under investigation.

In our second experiment queries were run on the first 5 minutes of each target image in our dataset, ranking content by mention frequency and removing all content that has appeared previously before the 5 minute target window. The ranked list of images was correlated to the target image expanded URI set to see how far up the ranking each target image came. The top 100 content items in each ranked list were also manually inspected to discover what percentage were eyewitness images and/or videos relating to the Paris shootings. The aim of this experiment is to examine if a combination of temporal segmentation and ranking could be used to support a real-time news feed for new unpublished eyewitness content and how much noise there might be for journalists to tackle.

These experiments were executed on a single machine (64bit Ubuntu 14.04LTS, Intel i7 8x2.7GHz, 16GB RAM). We used Python 2.7.6, PostgreSQL 9.3.10, Sesame 2.7.16 and GraphDB lite 6.1.

4.3 Experiment results

We show the results in figures 2 to 8 for each image based on the frequency of mentions for the extended set of URLs. The X-Axis shows the 10 minute segments within the first hour from the publication of the original image. The figures convey information about the popularity of the image, how and when it went viral and its credibility.

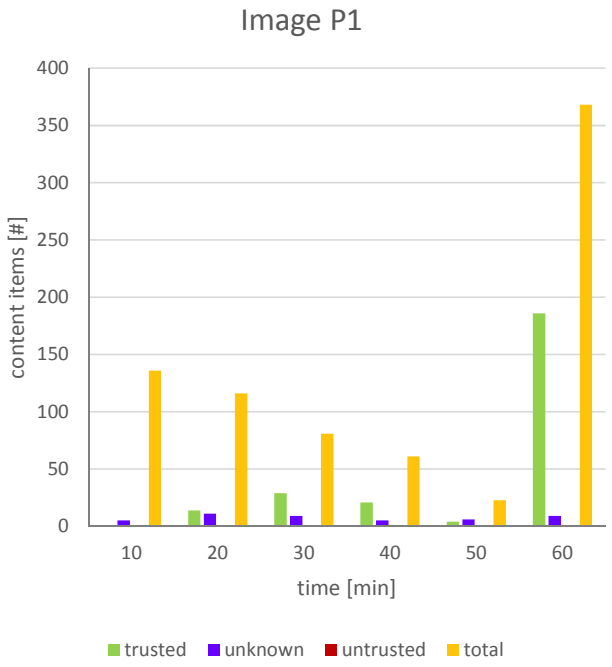

Figure 2. Number of tweets mentioning the URLs of image P1 in the first hour after publication, attributed to unknown and (un)trusted sources and the total mentions of URLs

[5] e.g. http://twitter.com/BBCBreaking/lists/news-sources/members

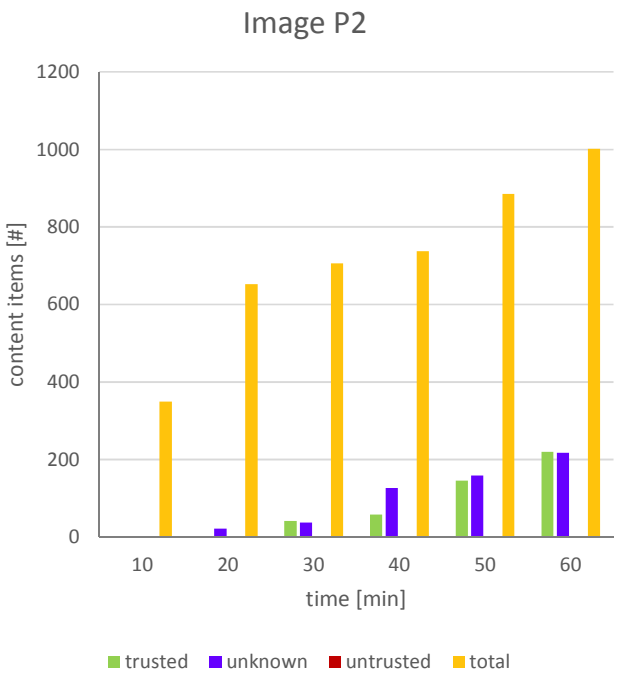

Figure 3. Number of tweets mentioning the URLs of image P2 in the first hour after publication, attributed to unknown and (un)trusted sources and the total mentions of URLs

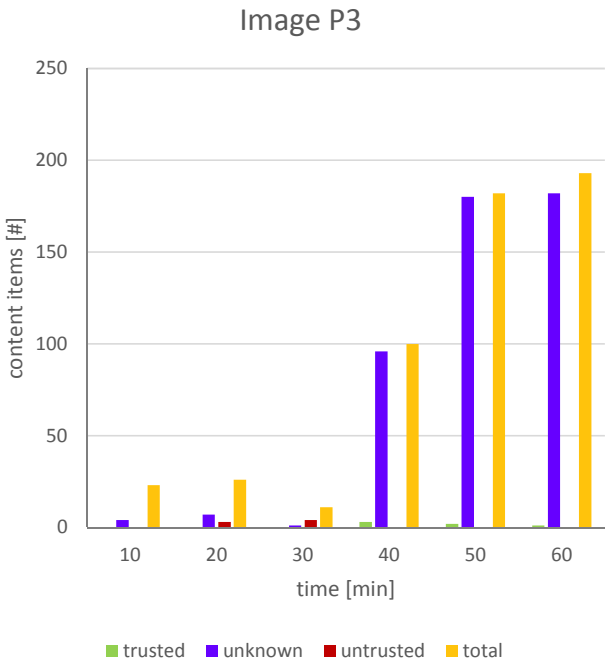

Figure 4. Number of tweets mentioning the URLs of image P3 in the first hour after publication, attributed to unknown and (un)trusted sources and the total mentions of URLs

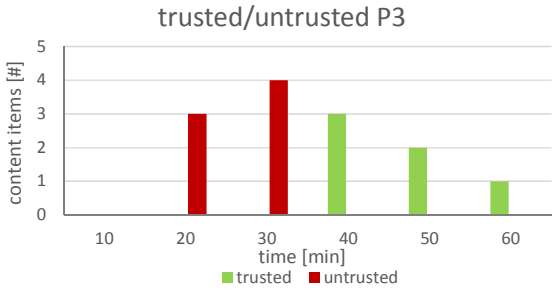

Figure 5. Detail: attribution to (un)trusted sources for P3.

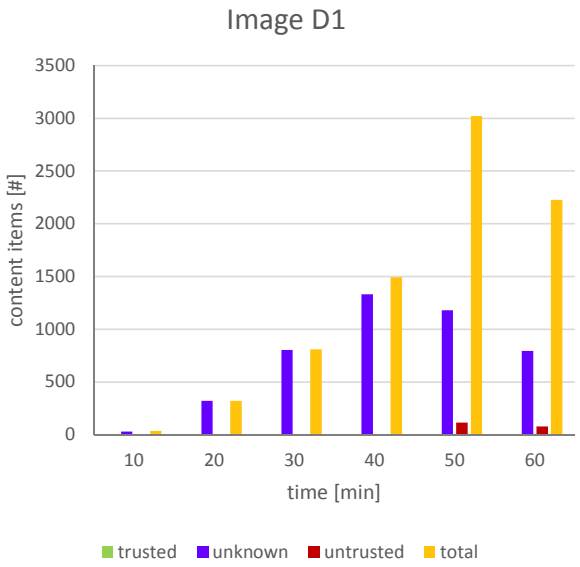

Figure 6. Number of tweets mentioning the URLs of image D1 in the first hour after publication, attributed to unknown and (un)trusted sources and the total mentions of URLs

Figure 7. Number of tweets mentioning the URLs of image P1 in the first hour after publication, attributed to unknown and (un)trusted sources and the total mentions of URLs

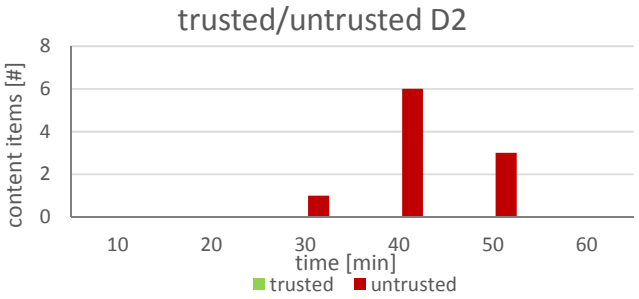

Figure 8. Detail: attribution to (un)trusted sources for D2.

Table 2. Overview of referenced URLs in content items within the first 5 min segment after the original tweet was published and how many of them contained both true and false eyewitness material (using manual inspection) in the first 100 URLs

Target Image	P1	P2	P3	D1	D2
ranking of target image set in total for 5 minute segment (top x percent)	9 / 653 (2%)	1 / 603 (1%)	61 / 1097 (6%)	427 / 11605 (4%)	1 / 11337 (1%)
total number of eyewitness content in 5 minute segment	25	2	12	29	30
unique number of eyewitness content in 5 minute segment	4	1	4	13	14

5. DISCUSSION

As shown in table 1, P1 comes from a civilian eyewitness. Apart from the tweet we investigated, he posted several other similar pictures of the shooting at the Paris bar "Le Carillon" as seen from his window. With relatively few followers, the image did not reach a bigger audience in the beginning. P2 has a higher volume and relates to the shootings at the café "Le Petit Cambodge". The author is a production assistant for a French TV channel and consequentially has more followers, resulting in a better spreading of the original image. P3 shows the café "La Belle Equipe" in Paris' Rue de Charonne with firefighters clearing the scene about 30 minutes after the shooting. The author of is again a civilian with few followers.

In image D1 the author is the official Twitter account of a US based news agency. "Herald de Paris" provides news in both English and French, drawing from journalists from external agencies around the world. Because it is ambiguously named, other users assumed it is a local newspaper. This assumption made it more credible than it actually is in this context, where it circulated a picture of the Eiffel Tower with the lights switched off. While locals know that this is the case every night at 1AM, the author made an effort to convince people it had happened in memory of the victims rather than to save electricity. With a high popularity, it managed to strike the chord of general sentiment in the chaos of the attacks. D2 is also a false rumour, relating to a gathering of people on the Place de la Republique, which took place after the Charlie Hebdo attacks in January 2015. It shows people holding up letters that spell out "NOT AFRAID". The author, a political journalist, has a large group of followers and knew about the impact an emotional message such as this would have on people around the world. The story got widely shared though the author never replied to verification questions asked by other journalists. It was debunked multiple times, with the first tweet only one minute later and a current photo showing the place empty 3 minutes after the original tweet was published.

P1, P2 and P3 are all images that turned out to be true. Figures 2 and 3 show how content items started to attribute the image to trusted sources from the second temporal segment onwards (i.e. after 10-20 minutes). The fact that the author of P1 posted multiple, very similar pictures which were shared within few minutes of each other accounts for the difference between the amount of shared URLs and the total amount of relevant content items. Hints concerning the veracity can be observed early on, with a peak at about 30min. The fact that this number falls between 30 and 50min does not affect the classification as an authentic image: the fact that it was shared by trusted sources at all is sufficient as an indicator. Judging from the falling amount of relevant content items it could have *not* gone viral. After about 50 minutes however, it does and it becomes clear that it is indeed authentic. The author of P2 being a professional causes the image to spread more gradually. Neither P1 nor P2 show occurrences of untrusted sources. P3 starts spreading after about 30 minutes (see figure 4). It is different to P1 and P2 because the image is picked up by untrusted sources first, see figure 5. After 30 minutes however, it became popular quickly. The reason for this is that from about 21:50, a link to a Mashable article was shared widely. Mashable links to the original tweet in its article. While very early on, the image was attributed to untrusted sources, presumably before being verified properly, trusted sources picked it up later, which suggests it has now been verified and is authentic after all.

For the false rumours D1 and D2, figures 6, 7 and 8 show the absence of any trusted source during the first 60 minutes. For image D1, the reason is that after only 12 minutes, the rumour was debunked, with multiple other debunking tweets following in the course of the early morning. However, as some people stated in their comments to the tweet, despite being untrue, people a) trusted the news magazine thinking it was local ("I think I can trust the local news outlet to know why the Eiffel Tower lights were out.", @laura_payton), and b) realised the image was fake (i.e. out of context) but thought it was a nice symbol ("Me too! but I decided to play along", @dbrabyn). This fact seems to have sufficed as a justification for less trustworthy news sources to continue sharing the image long after it had been debunked. For D2, figure 7 shows a decline in overall popularity only 20 minutes after publication. This is caused by the very early debunk and the good availability of historic images via TinEye. Both images were not attributed to any trusted sources but some untrusted sources appear in the figures 6 and 8, indicating that the tweet is a false rumour.

Overall, our experiments indicate that the veracity of a content item can be inferred after about 30 minutes by looking for reports by trusted sources, for real content, and untrusted sources for later debunked content. In general the trusted source metric correlates to the genuine nature of a content item, and untrusted source inversely correlates.

Table 2 shows the results of the second experiment. The total amount of URLs gives an indication as to how high the volume of content items was during the 5 minute period following the publication of the original content item. One possible reason why this number grows is that D1 and D2 were posted much later. By that time, more people were talking about the attacks. The position of the image URLs in the list of shared URLs gives an indication as to how popular the item was in the first 5 minutes. Considering the total amount of shared URLs, this result is promising. It places all of the selected content items in the top 1-6%, meaning a journalist scanning a social media stream for newsworthy content would not have to check hundreds or thousands of URLs but could focus on the top URLs. This does not mean all top URLs will

contain true images but they are more likely to be related eyewitness content. We analysed each of the images separately, filtering only URLs from the 5 minute segment immediately preceding the current segment. For each item, there is typically at least one duplicate item. This is because each image shared on Twitter has at least two URLs (the actual address of the image and a search-engine optimised version of it) plus the URL of the tweet that contains the image. In a commercial deployment duplicates could be quickly skipped over if presented visually to a user, or additional filters could be applied based on text and/or image pattern analysis to remove them automatically.

6. CONCLUSIONS

In our first experiment on the November 2015 Paris attack dataset we examined the velocity of breaking eyewitness content with respect to how (un)trusted attributed entities appear over time. For verification the central hypothesis of this work is that the "wisdom of the crowd" is usually no wisdom at all [6] and it is often better to base a decision on a single trusted voice than the noise in an "echo chamber" such as Twitter. Our results show that from about 30 minutes onwards verified reports for eyewitness content start to be published from trusted sources which could be used to assess the veracity of this content. If verification is required before this timescale then other methods are needed of course, such as traditional journalistic verification by attempting to contact the source directly and doing some factual cross-checking to show consistency and credibility.

Our second experiment examines the first 5 minutes of our breaking news events, which is of particular importance to some journalists. We show that temporally segmenting our data into 5 minute segments, filtering out content that has been seen previously and then statistically ranking by mention frequency is a promising way to filter content as it goes viral. We found that content ranked this way would have presented our ground truth images to a journalist in the top 6% of all trending URIs for each 5 minute segment. We thus think this approach is well suited to providing a real-time information feed and recommending possible new eyewitness content relevant to breaking news stories. This approach can also be combined with other state of the art filtering approaches (e.g. automated multimedia fake detection [2] [3] [5]) to further improve the quality of the recommended real-time content to journalists.

Under the REVEAL project we plan to extend our knowledge model to include other potentially relevant information used in the verification process. For example we are looking into cross-checking known facts about the weather and lighting conditions, obtained from dynamic lookup using known event timestamps and locations, with results from image weather classification algorithms. Furthermore we plan to conduct an ethnographic study with professional journalists to evaluate how they make verification decisions, what sources they cross-check and provide a ground truth dataset for the decisions made by our knowledge-based trust model.

Our goal is not to fully automate journalistic verification; all decisions ultimately lie with the journalist. Instead we want to support a semi-automated process, where focussed state of the art algorithms can assist journalists in filtering real-time content (i.e. boosting efficiency in identifying key content) and cross-checking facts against large volumes of social media and open data sources are not practical for humans to do under breaking news timescales. Such semi-automated approaches will free up journalists to focus their precious time on the more difficult and subjective verification tasks that benefit most from human attention.

7. ACKNOWLEDGMENTS

This work is part of the research and development in the REVEAL project (grant agreement 610928) supported by the 7th Framework Program of the European Commission.

8. REFERENCES

[1] Silverman, C. 2015. *Lies, Damn Lies, and Viral Content. How News Websites Spread (and Debunk) Online Rumors, Unverified Claims, And Misinformation*. Tow Center for Digital Journalism, Columbia Journalism School

[2] Boididou, C. Andreadou, K. Papadopoulos, S. Dang-Nguyen, D.T. Boato, G. Riegler, M. Kompatsiaris, Y. 2015. *Verifying multimedia use at mediaeval 2015*. Proceedings of the MediaEval 2015 Workshop, Wurzen, Germany, September 14-15, 2015

[3] Middleton, S.E. 2015. *Extracting Attributed Verification and Debunking Reports from Social Media: MediaEval-2015 Trust and Credibility Analysis of Image and Video*. Proceedings of the MediaEval 2015 Workshop, Wurzen, Germany, September 14-15, 2015

[4] Middleton, S.E. Middleton, L. Modafferi, S. 2014. *Real-Time Crisis Mapping of Natural Disasters Using Social Media*. Intelligent Systems, IEEE, vol.29, no.2, 9-17

[5] Middleton, S.E. Krivcovs, V. 2016. *Geoparsing and Geosemantics for Social Media: Spatio-Temporal Grounding of Content Propagating Rumours to support Trust and Veracity Analysis during Breaking News*. To appear in ACM Transactions on Information Systems: Special Issue on Trust and Veracity, ACM

[6] Hu, M. Liu, S. Wei, F. Wu, Y. Stasko, J. Ma, K. 2012. *Breaking News on Twitter*. CHI 2012, May 5–10, 2012, Austin, Texas, USA

[7] Schifferes, S. Newman, N. Thurman, N. Corney, D. Göker, A. Martin, C. 2014. *Identifying and Verifying News through Social Media*. Identifying and Verifying News through Social Media, Digital Journalism, 2:3, 406-418, DOI: 10.1080/21670811.2014.892747

[8] Vis, F. 2012. *Twitter As A Reporting Tool For Breaking News.* Digital Journalism, 1:1, 27-47, DOI: 10.1080/21670811.2012.741316

[9] Finn, S. Metaxas, P.T. Mustafaraj, E. O'Keefe, M. Tang, L. Tang, S. Zeng, L. 2014. *TRAILS: A System for Monitoring the Propagation of Rumors On Twitter*. Computation and Journalism Symposium, NYC, NY

[10] Zhao, J. Cao, N. Wen, Z. Song, Y. Lin, Y. Collins, C. 2014. *#FluxFlow: Visual Analysis of Anomalous Information Spreading on Social Media*. IEEE Transactions On Visualization And Computer Graphics, Vol. 20, No. 12