Cross-Media Linking in Times of Disaster

Gerhard Backfried
Sail Labs Technology
gerhard.backfried@sail-labs.com

Christian Schmidt
Sail Labs Technology
christian.schmidt@sail-labs.com

Gerald Quirchmayr
University of Vienna
Multimedia Information Systems Research Group
gerald.quirchmayr@univie.ac.at

ABSTRACT

Many possible links and connections can be observed between the different types of media used for communication during a crisis. These links can be detected and assembled to provide a more complete picture of events. They can be categorized according to the type of destination which yields important information for the gathering process as well as concerning general patterns of how platforms are connected. Tweets, posts and comments thus become parts of larger, linked sets of documents forming *compound-documents*. These documents stretch across media borders and platforms and provide context and broader information for individual entries. In the current paper we describe some of the links and linking behavior encountered during the floods in Central Europe of 2013 from the perspective of Twitter and Facebook.

Keywords

Crisis communication, data collection, natural disasters, situational awareness, Social Media

INTRODUCTION

Traditional Media, such as radio, TV or news-feeds have a long history of providing and disseminating information about crises and disasters. In certain areas of the world (and depending on a multitude of socio-economic, religious, cultural, educational etc. factors) these media may indeed still form the primary source of information for a substantial part of the population and stakeholders. Recent years have witnessed a dramatic surge in disaster-coverage on various Social Media platforms. Microblogging, social networking, photo- and video-sharing platforms are all being used by an increasingly large body of participants to document incidents, make sense of events and provide detailed insights. This increase has been facilitated by the rapid growth of portable devices and accompanied by the development and implementation of strategies on how to best combine such technological advances with humanitarian and crisis-management objectives (Chan, 2013; International Federation of Red-Cross and Red-Crescent Societies, 2013). The use of Social Media in reporting about crises and disasters frequently involves the (simultaneous) use of multiple platforms and modalities and often includes active linking (*mashing up*) of information within and between platforms as well as between traditional and social media. Whereas much work has been dedicated to the use of individual platforms, notably Twitter, relatively little work has been devoted to the investigation of the links created between

different types of media, the information presented by such links, their temporal and structural properties and the possibilities they offer. The current paper attempts to address this gap by providing some initial insights gained during a project focusing on cross-media and multi-media communication during crises and disasters.

EXISTING WORK

Social Media have been used for the documentation of events during disasters and crises for a considerable period of time. With the creation of platforms like Flickr and Facebook in 2004, YouTube in 2005, Twitter in 2006 or Instagram in 2010, these platforms also quickly became central elements in the coverage of disasters in the scope of citizen-driven information generation and -dissemination. First responders and stakeholders likewise quickly became aware of the benefits to be gained in terms of the possibilities of citizen involvement and for situational awareness and have embarked on assuming active roles on the same platforms. These activities and platforms have subsequently become the focus of research investigating their particular use during many real-world disaster-events. Even as early as 2005 during the London bombing attacks different types of media – photos posted on Flickr, textual content on Web-Sites of newspapers, coverage on TV stations and even entries in Wikipedia were already being mashed-up and connected by numerous actors (Peary, Shaw & Takeuchi, 2012). Several projects investigate the role of microblogging services, predominantly of Twitter or discuss the role Social Media platforms have played in the coverage of disasters (Hughes, 2009; American Red Cross, 2010; Holmes, 2011; Faustino, 2012; Nagar, 2012; DHS, 2013). (Liu, Palen, Sutton, Hughes and Vieweg, 2008) report on the role of the photo-sharing platform Flickr for six notable disasters between 2004 and 2007 and how cross-referencing of different media has made citizen journalism more prominent. More recently, the use of multiple (social media) platforms during an event, in particular the use of Twitter and Facebook during the floods in Colorado and in Germany in 2013 has been investigated by (St.Denis, Palen and Anderson, 2014) and (Kaufhold and Reuter, 2014). Based on experiences and interviews with digital volunteers (Reuter et al, 2015) propose a cross-social-media application for combination of social media sources to aid volunteers in disaster-response. We aim to take the cross-media approach one step further by not only including data from multiple social media platforms, but also by combining information from traditional and social media into a joint system for analysis. In (Backfried, Prinz, Göllner, Quirchmayr and Czech, 2014) a project aiming at providing such a broader approach and framework is introduced. Within this project, a corpus covering one particular disaster and spanning several media types has been collected. Based on this corpus, the current paper attempts to provide some quantitative insights into the cross-media linking behavior and patterns found in Social Media during this disaster (cross media linking between social media and traditional media will be addressed separately).

CORPUS

The floods in Central Europe during the summer of 2013 caused some of the worst flooding in over 500 years. Several countries were affected, among them Germany and Austria (Wikipedia). Several river systems were involved, most importantly the Danube and Elbe rivers and their tributaries. The events and the different languages spoken in the affected areas and countries resulted in active reporting and multi-lingual documents on all media channels and platforms in Europa and abroad.

The floods form one of the use-cases of a national Austrian security research project called QuOIMA (Backfried et al., 2014), focusing on the use of multi-media data and cross-media linking during crises and disasters. In order to allow for the analysis of communication and -patterns - in particular ones entailing multiple media - a corpus spanning different types of media and sources all pertaining to a single event has been created as part of the project. The data were collected during May and June of 2013 with the period of 20th of May to 23rd of June representing the core period of flooding for Austria (data collection was also carried out before and after these dates but has not been analyzed to the same level of detail as the focus of QuOIMA was primarily on Austria). Table 1 provides an overview of the different data types and amounts collected.

Medium	Amount	Comment
TV	218h	13 TV programs, 9 TV stations, 4 countries, German and English, news programs
Internet (Web)	61710 pages	102 sources, German, English
Twitter	470k tweets	German (mostly) English and Dutch
Facebook	9800 posts / comments	posts and comments from 16 public pages, German and English, from more than 1000 users
Press Agency	750 articles	Press-releases, German

Table 1. Overview of corpus

Data collection took place within a larger framework of data-gathering and –processing components, supporting several traditional as well as social media platforms. Content was collected using digital satellite (TV) feeds, RSS-feeds and various APIs. TV content was transcribed using automatic speech recognition. All web and social media content was processed by a chain of text-processing components. All collection activities took place in a continuous, real-time 7x24 manner.

LINKING

Many links or connections can be observed between the different types of media during a crisis: TV inserts list Twitter account names and hashtags, tweets link to web-sites of newspapers, newspapers list Facebook accounts and refer to TV programs. Many of these links can be detected and processed to arrive at a more connected picture of how information is being shared, assembling connections to provide a more complete picture of events. The linking behavior across platforms may display particular patterns, allowing to identify certain media to play more important roles during individual phases of a disaster or within a certain community. Likewise, these links can serve to automatically detect further sources allowing for the extension of an information gathering process. For instance, this way tweets become parts of larger, linked sets of documents, e.g. images which are linked-to and their associated meta-information, and form compound-documents stretching across media borders and platforms. In turn, these compound documents provide context and broader information for analysis and visualization.

In the current paper we describe some of the links and linking behavior encountered during the floods in Central Europe of 2013 from the perspective of Twitter and Facebook. The distribution of links across the two platforms, their temporal behavior and the effects on data collection strategies are addressed. Linking patterns beyond URLs, such as the use of common hashtags across platforms are expected to be present in the remaining parts of the corpus. These contribute further to compound-documents and are the subject of ongoing work.

TWITTER

Tweets were collected via the Twitter Streaming-API (Twitter) using a combination of accounts and hashtags. Collection was refined in several steps in an iterative manner. Hashtags were determined starting from a disaster-ontology expanding concepts to the actual hash-tags used during the particular disaster. Accounts were determined by including first responders, local and international agencies and individuals known to be active in times of disasters. This set was extended over time through analysis of communication patterns between users.

All URLs referenced in these tweets were expanded and the ones referenced 100 times or more and identifiable as belonging to one of the following classes (manually) assigned accordingly: *News Outlets* refer to TV-stations, newspapers, magazines etc. *Organizations* refers to public and non-public bodies, such as ministries, first-responders or political parties. *Weather/Disaster* refers to sites specifically reporting on weather, flood-level and disaster conditions.

The left section of Table 2 provides an overview of these categories of sites linked-to from tweets and the number of unique sites (domains) within each category (the right portion of the table displays the corresponding numbers for Facebook-posts and -comments).

Type	Twitter		Facebook	
	Count	Number (distinct)	Count	Number (distinct)
Social Media	25820	21	614	7
News outlets	41417	82	34	27
Organizations	4643	17	171	41
Weather/Disaster	13881	18	90	8
total	85761	138	909	83

Table 2. Linked-to sites for tweets, FB-posts and FB-comments

Table 3 lists the top 10 individual most popular linked-to sites. By far the most popular site is Facebook, where links typically reference individual images and photo-albums from the affected sites. Meta-data from these objects, such as location data or time-information, can be combined with meta-data from Twitter itself for confirmation and extension of information (Fuchs et al, 2013). By applying a comparison on dates, e.g. images of past floods (taken at identical sites) could be detected.

Figure 1 displays the distribution of linked-to sites according to their category (for sites referenced at least 100 times). Almost half of the links encountered in tweets link to traditional news outlets. An analysis of the origin of links shows that there is a great variety in the accounts linking to these sites, pointing to actual sharing of information which is considered important to a population rather than self-promotion of news-outlets. Social Media - and here foremost Facebook - are likewise linked-to frequently.

Domain	Count
www.facebook.com	10030
www.seismoblog.de	5703
youtube.com	4946
instagram.com	3524
www.spiegel.de	3187
www.welt.de	2552
www.focus.de	2263
www.mz-web.de	2170
www.tagesschau.de	1778
www.n24.de	1717

Table 3. The 10 most frequently linked-to sites from tweets

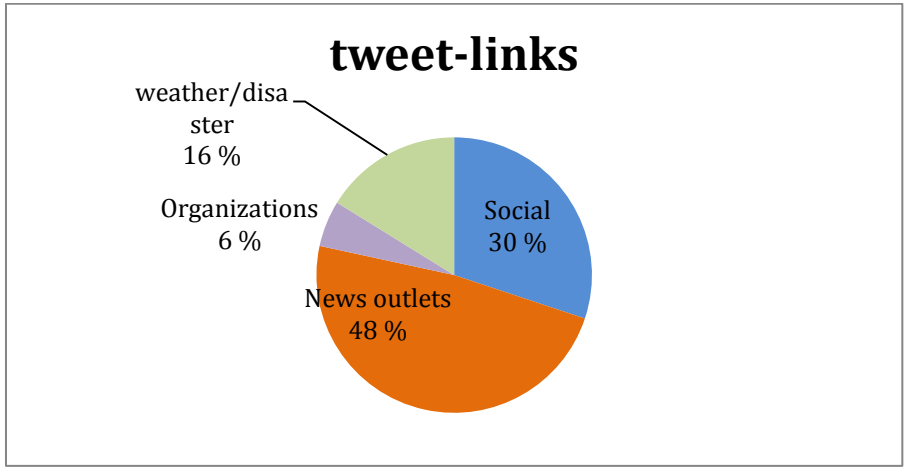

Figure 1. Distribution of links from tweets across categories

An analysis of the behavior of these links over time indicates that there is no clear trend towards linking to one type of category, but rather that all types of media are linked-to during all stages of a disaster in a similar manner.

Bot-detection was applied to all users producing tweets similar to (Chu et al, 2012). The linking behavior (whether URLs were contained) was observed to be different, with links in human-generated tweets following the general trend and volume of tweets over time and bot-generated tweets remaining relatively stable. However, as depicted in figure 2, also for bots a slight increase can be observed which seems to indicate that bots might have been created or activated during the peak period of the flood. Domains of links included in tweets by bots are highly homogeneous which provides a further parameter for automatic bot-detection.

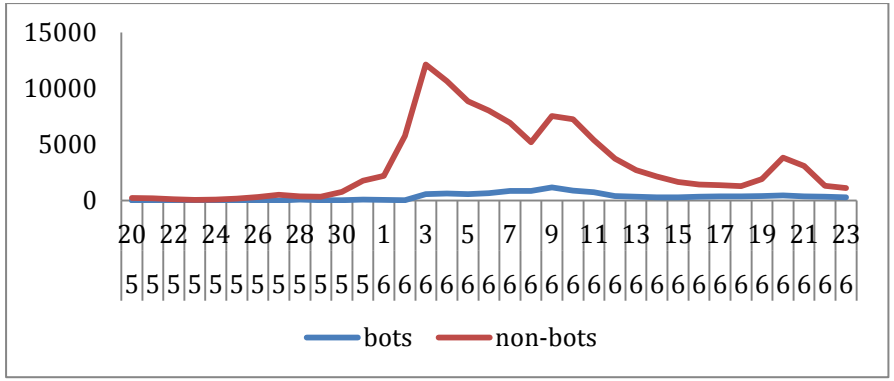

Figure 2. Distribution of links from tweets generated by bots and humans

FACEBOOK

The posts and associated comments on 16 public Facebook pages were collected using the Facebook Graph-API (Facebook). The set of public pages comprises individuals and organizations and was refined iteratively before and during the floods.

All URLs referenced in posts and comments were expanded and classified in analogy to the ones encountered in tweets. The right section of Table 2 provides an overview of the categories which are linked-to, table 4 lists the 10 most frequent domains.

domain	count
www.facebook.com	592
www.gdacs.org	52
www.thw.de	50
www.roteskreuz.at	25
www.skywarn.at	33
www.youtube.com	14
www.drk.de	11
www.helpstars.at	10
www2.land-oberoesterreich.gv.at	9
oxan.to	9

Table 4. 10 most frequently linked-to sites from FB posts and comments

Similar to the sites linked-to from tweets, Facebook is the most popular site referenced from posts and comments (highly influenced by the fact that the link origin is Facebook). As for tweets, the second most popular Social Media site is YouTube, providing clips of flooded areas and relief efforts. The variety of Social Media platforms seems to be lower than on Twitter. Unlike with tweets, image sharing sites, such as Instagram or Flickr are not referenced at all. Whether this is a general tendency – having an effect on potential data collection strategies – or an artefact of the corpus and/or the behavior of the affected population will have to be cross-checked against other corpora.

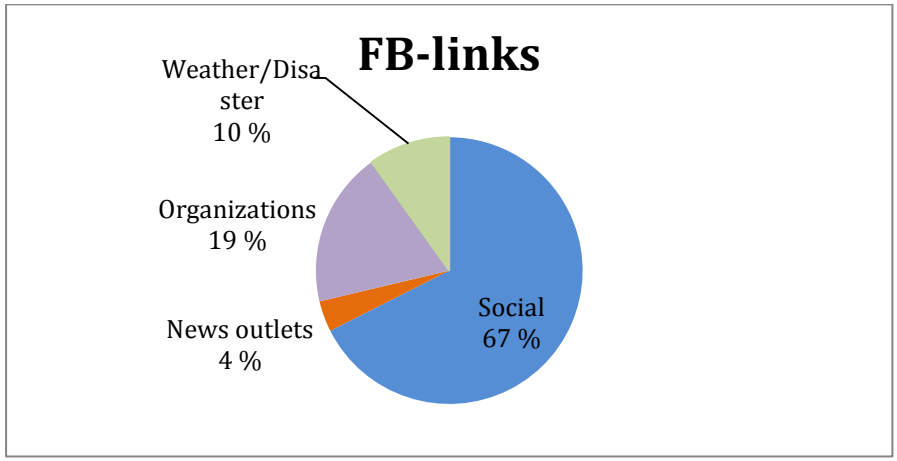

Figure 3. Distribution of links from FB posts and comments across categories

In contrast to Twitter, Social Media are by far the most frequently linked-to category. Links to the sites of organizations such as the Red Cross, THW or a provincial governmental site are more common than on Twitter. This may also point to the fact that organizations such as the Red Cross may be more active on Facebook and that users on Twitter are generally not aware of content provided by these organizations and consequently do not link to them. News outlets are referenced in a mere 4% of all cases and do not appear in the first 10 positions of popular links. This may indicate that content typically covered by news-sites might also be provided within the platform itself, at least on the observed platforms. Figure 3 displays the distribution of linked-to sites according to their category for all links encountered in the collected posts and comments.

CONCLUSIONS

Due to the different levels of popularity and activity, the absolute numbers of links encountered are difficult to compare. However, some commonalities as well as differences seem to be detectable. Links from both sites, Twitter and Facebook, frequently reference content from other sites. The two platforms differ in the categories and distribution of sites which are linked-to, but are behave similarly when observed over time: all links from platforms follow the general trend of volume of posts with the exception of bot-generated tweets as noted above. Facebook and YouTube are among the most popular link destinations from both platforms and thus form logical additions in terms of data-collection, analysis and visualization activities. Image-sharing sites were only found to play a significant role for tweets. This could point to a principally different linking behavior concerning photo-sharing on the two platforms. However, this finding still needs to be confirmed by examining other corpora. Links to traditional news-outlets seem to be more frequent on Twitter, pointing to the fact that relevant content may be perceived to be available within the platform itself for users on Facebook. Active self-referencing of content within Facebook by organizations such as the Red Cross might also play a strong role in this respect. Many news-outlets are active on Twitter and use tweets to provide headlines and links to their own Web content. The diversity of users linking to this content seems to indicate that the linking frequency is not due to self-promotion but rather to genuine interest of many users. Differing underlying strategies and assumptions on quality-assured information and engagement of readers for feedback and comments might likewise play a role. Regarding the extension of collection activities and discovery of further sources, the different linking behavior between Twitter and Facebook should be taken into account when using these platforms as drivers for source discovery. Weather specific sites are referenced from both platforms with Twitter exhibiting a slightly higher percentage and greater variety. When observing the temporal behavior of links, no difference in the distribution and different media sites could be found. The occurrence of URLs within tweets and posts follows the general volume of activity, which seems to indicate that all media should be taken into account the same way when collecting information during all stages of a disaster.

FUTURE WORK

Whereas the current work only examined links emanating from Social Media, links emanating from traditional media also exist and should be taken into account. The remaining sections of the corpus are expected to contain further links, which merit exploration and whose behavior is expected to influence the collection and processing strategies. E.g. through the application of OCR on TV images or via automatic speech recognition, hashtags might be discovered, leading to connections between TV programs and Twitter. Sharing of hashtags between image-sharing sites, Facebook and Twitter might establish further links between these platforms. The notion of compound documents will thus be extended and include these links as well.

The authors are aware that the findings on this particular corpus are intimately tied to the particular event, the population of Social Media users and involved organizations. This behavior might differ by region and is likely to also change over time as stakeholders adapt their strategies on increase their presence on individual platforms. Corpora collected during other natural disasters will thus need to be examined with regard to the linking and compared to the results obtained on the 2013 floods.

ACKNOWLEDGMENTS

This work was supported by the FFG project QuOIMA (836278) under the Austrian National Research Development Programme KIRAS.

REFERENCES

1. American Red Cross (2010), The Case for Integrating Crisis Response with Social Media, White Paper, http://bit.ly/1vSabcX on 2015/02/28
2. Backfried, G, Prinz, K., Göllner, J., Quirchmayr, G., Czech, G., (2014) Cross-media, Multimedia, Multilingual Communication in Open Sources During Crises and Disasters, ISCRAM 2014, University Park , PA, USA
3. Chan, J.C., The Role of Social Media in Crisis Preparedness, Response and Recovery, VANGUARD Report, http://bit.ly/191xdTk on 2015/01/30
4. Chu, Z., Gianvecchio, S., Wang, H., Jajodia, S., (2012) Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?, IEEE Transactions on Dependable and Secure Computing, 11/2012; 9(6):811-824. DOI: 10.1109/TDSC.2012.75
5. DHS Report, Lessons Learned: Social Media and Hurricane Sandy, from http://1.usa.gov/1DuKtsA, 08/30.2013
6. Facebook, The Graph API, https://developers.facebook.com/docs/graph-api on 2015/01/30
7. Faustino, J.D., Liu, B., Jin, Y., (2012) Social Media During Disasters: A Review of the Knowledge Base and Gaps, Final Report to Human Factors/Behavioral Sciences Division, U.S. Department of Homeland Security, College Park, USA, 2012
8. Fuchs, G., Andrienko, N., Andrienko, G., Bothe, S. & Stange, H., (2013) Tracing the German Centennial Floods in the Stream of Tweets: First Lessons Learned, SIGSPATIAL Intl Workshop on Crowdsourced and Volunteered Geographic Information, 2013, Orlando, FL, USA
9. Holmes, W., (2011) Crisis Communication and Social Media:A dvantages, Disadvantages and Best Practices, Univ. of Tennessee, CCISymposium, 2011, Knoxville, TN, USA
10. International Federation of Red Cross and Red Crescent Societies, (2013) World Disasters Report 2013, http://bit.ly/1EFbVbj on 2015/02/28
11. Kaufhold, A, Reuter, Ch., (2014) Vernetzte Selbsthilfe in Sozialen Medien am Beispiel des Hochwassers 2013, (Linked Self-Help in Social Media using the example of the German floods of 2013), i-com Zeitschrift für interaktive und kooperative Medien, 13(1), 20-28, 2014
12. Hughes, A. L., Palen, L., (2009) Twitter Adoption and Use in Mass Convergence and Emergency Events, ISCRAM 2009, Gothenburg, Sweden
13. Nagar, S., Aaditeshwar, S., Joshi, A., (2012) Characterization of Social Media Response to Natural Disasters, SWDM 2012, Workshop, Lyon, France
14. Peary, B., Shaw, R. and Takeuchi, Y., (2012) Utilization of Social Media in the East Japan Earthquake and Tsunami and its Effectiveness, Journal of

Natural Disaster Science, Volume 34, Number 1, 2012, pp3-18

15. Reuter, Ch., Ludwig, T., Kaufhold, M.A., Pipek, V. (2015) XHELP: Design of a Cross-Platform Social-Media Application to Support Volunteer Moderators in Disasters. 33rd International Conference on Human Factors in Computing Systems (CHI '15), Seoul, Korea
16. Sophia B. Liu, Palen, L., Sutton, J., Hughes, A. L., Vieweg, S., (2008) In Search of the Bigger Picture: The Emergent Role of On-Line Photo Sharing in Times of Disaster, ISCRAM 2008, Washington, D.C., USA
17. St.Denis, L.A, Palen, L., Anderson, K.M, (2014) Mastering Social Media: An Analysis of Jefferson County's Communications during the 2013 Colorado Floods, ISCRAM 2014, , University Park , PA, USA
18. Twitter, The Streaming APIs, https://dev.twitter.com/streaming/overview on 2015/01/30
19. Wikipedia, http://en.wikipedia.org/wiki/2013_European_floods available on 2015/01/30