

Stuttgart's black Thursday on Twitter: Mapping Political Protests with Social Media Data

Andreas Jungherr, MA
Research Associate, Chair for Political Sociology
Otto-Friedrich-Universität, Bamberg, Germany
E-Mail: andreas.jungherr@gmail.com

Pascal Jürgens, MA
Research Associate, Department of Mass Communication
Johannes Gutenberg-Universität, Mainz, Germany
E-Mail: pascal.juergens@googlemail.com

Analyzing Social Media Data and Web Networks: New Methods for Political Science, ed. Rachel Gibson, Marta Cantijoch, and Stephen Ward. New York, NY: Palgrave Macmillan (Forthcoming).

Draft
v.3.0
(17. March 2013)

This is a preprint. Please cite the published version of this paper.

Abstract

The night of Thursday September 30, 2010 to Friday October 1, 2010 brought one of the heaviest clashes between protesters and police in Germany's recent history. The protesters opposed plans to demolish a train station in the town of Stuttgart, Baden-Württemberg, and to replace it by an underground station. The events on September 30 were triggered when construction workers started cutting down trees under police protection. At the end of the night roughly 400 protesters had been injured and the events had sent shockwaves through Germany. One of the communication channels protesters used was the microblogging service Twitter. Protesters and supporters not present in Stuttgart used Twitter messages marked by the hashtag #s21 to exchange news, links to media content, and links to audio and video live streams. This quickly increased the visibility of the events well beyond Stuttgart while they were still unfolding. We use Twitter messages of Germany's 80.000 most prominent Twitter users to develop a timeline of the events of that night. We analyse this data (Twitter messages posted on 30 September and 1 October 2010 containing the hashtag #s21) with four distinct approaches for event detection: 1. Local maxima in the total volume of messages containing the hashtag #s21; 2. First occurrences of messages that were retweeted most often; 3. First occurrences of links that were posted most often; 4. Thresholds in the relative frequency of word stems used in messages. We are thus able to identify various discrete steps of the protest, its buildup and its aftermath. Also we are able to compare the results produced by these different methods. This paper thus illustrates the potential of social media for event detection based on bursty patterns in textual data.

Andreas Jungherr is a research fellow at the chair of political sociology at the Otto-Friedrich-Universität, Bamberg, Germany. He is co-author of the book *Das Internet in Wahlkämpfen: Konzepte, Wirkungen und Kampagnenfunktionen* (Wiesbaden: Springer VS). His articles have appeared in *Social Science Computer Review*, *German Politics, Policy & Internet*, *Zeitschrift für Parlamentsfragen*, *Internationale Politik* and *Transformative Works and Cultures*. He may be contacted at andreas.jungherr@gmail.com.

Pascal Jürgens is a research fellow at the department of mass communication at the Johannes Gutenberg-Universität, Mainz, Germany. His research interests include political communication, social networks and fragmentation phenomena on the internet. A second focal point lies in the development of methods suited for the study of digital behavioral data. He may be contacted at pascal.juergens@googlemail.com.

INTRODUCTION

Event detection based on textual data is an approach often used in the social sciences. The method has been used predominantly in the fields of international politics (Schrodt, 2010) and public opinion research (Landmann and Zuell, 2008). Event detection presupposes that major events leave traces in textual documents. By automatically identifying events in publicly available documents researchers can establish timelines of events relevant to their research. For example, in international politics, researchers work on how to reliably identify political actors, time, and topics from official documents, hoping to establish comprehensive and detailed maps of international treaties and conflicts. Based on these maps they aim to develop models of the dynamics of conflict (Brandt, Freeman and Schrodt, 2011). In public opinion research one goal is to automatically deduce major events from newspaper coverage. This might be a first step in calculating the impact of these events on changes in public opinion (Landmann and Zuell, 2008).

Most research in this area has focused on event detection based on textual data that filter events through structured reports, be it official documents, or newspaper articles (Allan, 2002; Kleinberg, 2003). This has the benefit that researchers are able to analyse a textual corpus focusing on relevant aspects of an event. The authors of these documents (that is officials, journalists) edited these texts consciously so that they contain relevant information. Thus researchers focusing on these documents potentially find a high signal to noise ratio (that is relevant information to irrelevant information) in these documents. But exactly the process of filtering relevant information by authors removes these documents one step from the actual events themselves. Official documents or newspaper articles often offer a summary of relevant actors, events, or outcomes of a topic under investigation. They are after

the fact accounts, not observations of unfolding events. Thus potentially relevant steps of the event might be missing in these accounts and remain hidden in analyses based on them. For a researcher interested in regular dynamics of conflicts and treaties this might seem a reasonable trade-off, but for those interested in the dynamics of protest, the chain of micro-events that constitute a protest event might hold important meaning. Clearly the analysis of textual data closer to the events of interest holds potential for social scientists. The ever-growing adoption of social media services provides researchers with data of that kind.

Increasingly people use social media services to document their lives, comment on events, or communicate with each other. While this activity can come in many forms (for example a user might take a photo of a protester being carried away by the police and directly post it on a photosharing service, or instead she might use her mobile phone to film the incident and post it on a videosharing site) but most of it will come in form of time coded textual status updates that lend themselves to computer assisted analysis (for example a user writes a short update on her Twitter feed that the police is carrying protesters away). Analysing these data offers researchers a closer look at the steps that constitute an event (for example a protest). Unfortunately this benefit is offset by the noise of unrelated information that surrounds the information of interest. Most social media users do not attempt to document events impartially as they unfold. Most users post updates on mundane details of their lives. They are not necessarily journalists but might be passers-by or participants in social events. Still they might document parts of these events on social media channels. Thus unintentionally each user becomes a sensor of her surroundings. The challenge for researchers attempting to use social media data to

document socially relevant events is to cut through the noise of unrelated information and identify those pieces of text that hold meaning.

In this paper we show that the analysis of messages published on the microblogging service Twitter can be used to establish a timeline for political events. We analyse Twitter messages by the 80.000 most prominent Twitter users in Germany. In our analysis we focus only on messages commenting on the highly contentious protest against the controversial project ‘Stuttgart 21’ in the night of September 30 going on October 1, 2010. ‘Stuttgart 21’ provides us with a case study that shows the potential of event detection with data collected from social media services. We identified relevant tweets by their use of the hashtag #s21. On September 30 and October 1, 2010 46.789 Twitter messages containing the hashtag #s21 were posted by 7.793 of the Twitter users in our sample. In our analysis the protests of that night reacting to the project ‘Stuttgart 21’ become the event. We are trying to identify the steps contributing to this event by the analysis of the 46.789 Twitter messages containing the hashtag #s21.

For our analysis we use four different approaches to event detection and compare their results. These approaches look for local maxima in the total volume of messages, the first occurrences of messages that were highly retweetet, the first occurrences of URLs on the Internet that were highly linked to and finally by the examination of words that were only prominent during specific time intervals of the protest (an approach originally proposed by Shamma, Kennedy and Churchill, 2011). We show that the microblogging service Twitter is a valuable tool for the mapping of political events.

TWITTER AS DATA SOURCE FOR EVENT DETECTION

The growing use of online tools and social media services has provided companies and researchers with an ever-increasing amount of rich data on human behaviour. Specifically data collected on the microblogging service Twitter (<http://twitter.com>) has become the focus of various research projects.

Twitter enables users to post short text messages (up to 140 characters in length) on personalised profiles. These Twitter feeds and the messages posted on them have URLs and are publicly accessible. The exception are cases in which users explicitly state that their feed is private and thus only accessible to users previously approved by them. Twitter users are able to subscribe to other Twitter-feeds to regularly receive updates. Thus each Twitter account is connected to accounts of users whose owner subscribed to (in Twitter terms ‘following’) and the accounts of users who decided to subscribe to it (‘followers’). The limit of 140 characters per message led to the widespread adoption of usage conventions in which regularly used abbreviations help to discern meaning. If users want to post a Twitter message on a given topic they use a keyword or commonly agreed upon abbreviation and precede it with a ‘#’ (hashtag, for example tweets commenting on the project ‘Stuttgart 21’ were marked by the hashtag #s21). This convention helps researchers to automatically identify tweets reacting to specific events, commenting on topics, or adding to a meme. If a user chooses to write a public message directed to another user she can do so by preceding the message with an ‘@’ followed by the username of the addressed person (for example a public message addressed to one of the authors of this paper would be preceded by @ajungherr or @pascal). If a user reads a tweet which she thinks important or witty and wants to bring it to the attention of her followers, she can do so by retweeting it. She can do this by copying the

original message preceded by the abbreviation ‘RT’—for retweet—followed by ‘@’ and the username of the original author. These two conventions—@message and retweet—enable researchers to extract social networks formed by communication activities by users. This is a powerful addition to the examination of follower/following networks of Twitter users.

Researchers are able to access Twitter’s data with various approaches. This paper cannot offer a systematic overview on different approaches to collect data from Twitter, but two approaches seem to dominate the relevant literature. One approach is to use Twitter’s API (application programming interface). An API provides outsiders with standardised access to a service’s databases. The API provides researchers with the message, the username of its author, a unique time stamp, the name of the third-party service the message was posted with, and the location where the tweet was sent from (provided a user enabled the geolocation option). Twitter started out with a relatively open data access policy that allowed users to run up to 2000 queries per hour on Twitter’s search API. This type of access is no longer provided; instead there are stricter (but unspecified) numbers of queries one can run on Twitter’s search API. In addition to this, Twitter offers access to random samples of the total stream of Twitter messages that provide users with a fixed percentage of the total amount of messages posted. It is difficult to determine the data quality provided by the Twitter API. There are indicators that the Twitter API provides researchers with systematically divergent data, dependent on whether researchers used search queries or accessed the sample stream. If unacknowledged, these differences can lead to biased results (González-Bailón et al, 2012). Without access to Twitter’s infrastructure, the precise nature of the sampling algorithm cannot be verified. Our dataset is based on a now defunct sampling mechanism which Twitter

describes as merely selecting the first N of 100 Tweets (White et al. 2012). The ID values of tweets used to increase linearly. Twitter's sampling algorithm then selected messages by calculating the modulus and returning tweets with certain remainder values, depending on the user's access level. All in all we are fairly confident that there is no significant impact of the sampling methodology on our dataset, especially since we only used it in order to bootstrap our own sample of German users. The messages of these users were collected independently of Twitter's random messages sample (see below).

Another approach is the use of third party applications that collect data on Twitter (for example by using the Twitter API, by scraping Twitter's openly accessible websites, etcetera). These applications offer researchers ease of use but potentially introduce a new black box in the data acquisition process. Still, tools like DiscoverText (<http://discovertext.com>) or yourTwapperKeeper (<https://github.com/jobrieniii/yourTwapperKeeper>) (Bruns and Liang, 2012) are becoming increasingly popular among researchers. A systematic comparison of data provided by these services and the Twitter API remains to be done to access the potentials and problems associated with each approach. We collected the data for this paper by using Twitter's streaming API.

Research focusing on Twitter could be grouped in three approaches: research interested in specific usage practices and the adoption of Twitter in various communities (for example Crawford, 2009; Marwick and Boyd, 2011); research interested in network structures on Twitter and information flows through these networks (for example Cha et al, 2010; Jürgens, Jungherr and Schoen, 2011); and research interested in using Twitter data to analyse or predict human behaviour and events offline (for example Asur and Huberman, 2010; Chew and Eysenbach, 2010;

Gayo-Avello, Metaxas and Mustafaraj, 2011; Jungherr and Jürgens, 2013). This paper clearly falls in the third group.

The idea behind research using Twitter data to detect major events is that each Twitter user is a sensor that documents her observations of reality in her messages. While most of these messages might document mundane details of her daily activities, others might address a social event the user might participate in (for example a sport event watched on TV) or an event she accidentally witnessed (for example police action during a protest). For events with high popular appeal (for example TV-shows, or the death of a celebrity) or social relevance (for example political protests, or natural disasters) it is reasonable to assume that many Twitter users tweet their reactions or observations. In the process of formulating their individual observations of the unfolding events they necessarily code their subjective impressions in a common vocabulary. This makes them automatically identifiable as signals referring to the same object. The sudden increase in messages on a certain event or topic produces automatically discernible patterns since these messages typically share attributes in semantic structure, their vocabulary, the use of hashtags, time stamps or linked content. Thus social events leave an imprint in Twitter data through clearly identifiable clusters of similar messages, which in turn might be automatically detected.

Various research communities have approached event detection with Twitter data with different aims. Some researchers try to detect potentially catastrophic events as they are unfolding and thus use Twitter as an early warning system (Sakaki, Okazaki and Matsuo, 2010) or to increase situational awareness in emergencies or humanitarian missions (Verma et al, 2011). Researchers also tried to use Twitter messages to determine the structure of big broadcast events based on the dynamic

and persistence of spikes in the use of event specific terms (Shamma, Kennedy and Churchill, 2009; Shamma, Kennedy and Churchill, 2011). Other researchers work on event detection algorithms in the hope of improving real-time search results with Twitter data (Becker, Naaman and Gravano, 2011; Chakrabarti and Punera, 2011; Petrovic, Osborne and Lavrenko, 2010; Weng and Lee, 2011).

The obvious potential of Twitter as a data source on human behaviour and interests should not blind us to the fact that Twitter's user base is still comparatively small and far from representative (Smith and Brenner, 2012). Attempting to draw conclusions on behaviours or interests of the population of a given country based merely on data produced by Twitter users of that country seems highly optimistic (Jungherr, Jürgens and Schoen, 2012). So far only a few studies have looked at the specific socio demographic composition of Twitter users. Their results suggest that Twitter users in a given country are —at least at this point in the adoption process— far from representative of other Internet users and the population as a whole (Busemann and Gscheidle, 2012; Smith and Brenner, 2012). This does not invalidate research based on Twitter data, but it means that researchers have to pay special attention to the interpretation of their results.

For the purposes of this paper we are interested in whether Twitter data allow the automated mapping of events during the unfolding of a political protest. This seems a sensible proposition since Twitter has become a very popular tool for users to comment on politicians, campaigns or political events (Australia: Bruns and Burgess, 2011; Germany: Jürgens and Jungherr, 2011; Jürgens and Jungherr, forthcoming; Netherlands: Vergeer, Hermans and Sams, 2011; Political protests: Segerberg and Bennett, 2011; Spain: González-Bailón et al, 2011; UK: Jackson and Lilleker, 2011; USA: Smith, 2011). For this paper the non-representativeness of

Twitter users is not an issue. We use Twitter data to examine if patterns in messages addressing the protests against ‘Stuttgart 21’ (#s21) correspond to offline events. To answer our question we do not need representativeness, we need a high volume of messages. Our analysis becomes possible since ‘Stuttgart 21’ —as we will show— generated interest among German Twitter users. They commented on the events as they were unfolding. This is positive but does not have to be true for other political events.

FOUR APPROACHES TO EVENT DETECTION WITH TWITTER

In this chapter we look at data documenting all Twitter messages containing the hashtag #s21 on September 30 and October 1, 2010 when major protests took place against ‘Stuttgart 21’. We then examine whether different analytical approaches show patterns that correspond with the occurrence of discrete developments in the actual protests. Our main objective is to examine the potential benefits and limits of different approaches to event detection using Twitter data. It is important to note that we use data on an event that happened in the past. Our analytical approaches can rely on the fact that data patterns at any given point of our analysis can be compared to data patterns at all other points during the time span of interest. This facilitates the analysis. There are other attempts by researchers that use different approaches to event detection in real time (for example Chakrabarti and Punera, 2011; Nikolov, 2012). To us this seems motivated less by the attempt to determine the structure of offline events through patterns in online data but more by using online data to determine the most important topics of online buzz at any given moment. The question we are addressing in this paper is not ‘how can we accurately

measure or predict levels of online buzz?’ but ‘is it possible to detect meaningful events based on the analysis of online data, specifically Twitter?’. For our goals the use of data sets documenting discrete events in the past seems unproblematic.

In this paper we compare the results of four approaches to event detection with Twitter data: local maxima in the total volume of messages, the first occurrences of messages that were highly retweeted, the first occurrences of URLs that were highly linked to and finally by the examination of words that were only prominent during specific time intervals of the protest.

1. Volume: this approach is solely concerned with the volume of messages and local maxima. This follows a simple assumption: the more users talk about a topic (measured by hashtags), the more important that topic is. By extension, the more they talk about it at a certain point in time, the more important or salient the topic is at this particular moment. While this approach is rather simplistic, there are still valid inferences to be drawn from it. For example, the mere fact that messages using a given hashtag follow a distinct pattern can often be directly interpreted. A sudden rise in tweets will signify rising interest and potentially point to a cause that at first glance might remain invisible to the analyst.

2. RTs: this approach focuses only on the occurrence of those tweets that were reposted (retweeted) the most. The premise is that users will select and redistribute tweets with especially high informative value or of high novelty value. These tweets might potentially refer to key events during the protest.

3. URLs: another approach using the same logic focuses on the most popular URLs that were linked to in messages containing the hashtag #s21. This approach offers further information as it might be that digital traces found on Twitter only echo existing reports by established mass media. Thus event detection with social media

data would be redundant to event detection based on news reports. If this were true, we expect two observations in the results provided by URLs: (1) time stamps indicated by the first occurrences of popular URLs should be delayed in comparison to the actual protests and possibly also in comparison to the time stamps as provided by local maxima in volume; and (2) most of the salient URLs should link to web pages of established media.

4. Peakiness: this term describes an approach introduced by Shamma, Kennedy and Churchill (2011) that offers a less simplistic approach to event detection than the former. Shamma, Kennedy and Churchill introduce the ‘peakiness’ value as the number of word occurrences within a time window divided by the number of occurrences within the entire reference time span. The value (ranging from zero to one) denotes how densely the use of a word is ‘lumped together’ in time. A peakiness of .5 means that half of the total uses of a word appear within one time window. Thus it is possible to identify the appearance of new and rare words during a short time span. In our case we can expect that words are very peaky if a clearly named object or action is only refereed to during a discrete time span in the complete run of the protests. At the same time, terms which are mentioned more or less constantly are spread out during the whole time span and hence not peaky. The substantial benefit for the detection of discrete steps in an event is that even keywords with a clear relation to the protests are filtered out if they are used ubiquitously. As we will show, there are several of these steps that can be successfully identified through the peaky characteristics of words referring to them. While the concept of peakiness has advantages, its general applicability remains to be shown. The use of this approach forces researchers to make choices in order to get promising results with their respective data sets. Most notably, researchers will

want to set a minimum peakiness threshold for word stems indicating discrete steps of an event in question. Secondly, the time window for which the peakiness value of a word is calculated has to be chosen carefully: if too small, there may not be any term whose mentions are packed enough to fit into one window, so peakiness will be low overall. If the time window is too long, not much insight into an event can be gained. Additionally, the value obviously depends on the total length and volume of a sample. This means that different datasets can only be compared if some sort of normalisation is used.

We calculated the peakiness of word stems (for example ‘Baum’ and its plural, ‘Bäume’ were counted as multiple occurrences of the same word), hashtags and URLs. We systematically varied the time window for our analysis between one to four hours and compared peakiness results. We found that for our time span of two days and the nature of the events documented by our data, the most relevant results were obtained for time windows of one hour.

STUTTGART 21

One of Germany’s most contentiously discussed topics in 2010 was an infrastructure project in Stuttgart called ‘Stuttgart 21’. Stuttgart is a town in the southwest of Germany and the state capital of Baden-Württemberg. ‘Stuttgart 21’ (#s21) is an infrastructure project with the plan to move Stuttgart’s central train station underground to increase its transit capacity. Since its inception in the early 1990s the project has met with strong resistance that reached its zenith in the second half of 2010 with regular demonstrations attended by participants in the tens of thousands. The protests attracted massive attention on various social media channels by both protesters and supporters. This makes ‘Stuttgart 21’ a promising object to

test the mapping of campaigns based on social media data (for a comprehensive discussion of the protests see Gabriel, Schoen and Faden-Kuhne, 2013).

In the night of September 30 to October 1, 2010 the protests escalated. Under police protection construction workers started cutting down trees. This led to heavy clashes between police and protesters during which up to 400 people were injured (sueddeutsche.de, 2010). The following day, in reaction to the clashes of the night before, between 50.000 and 100.000 (the sources vary) protesters took to the streets (Ternieden, 2010). Shortly after the night, that day became known as Black Thursday (Bilger and Raidt, 2011). Messages posted on Twitter during this night serve as the basis of our analysis.

Both protesters and supporters of 'Stuttgart 21' relied heavily on social media tools for organisation and visibility during the protests (Jakat, 2010; Mader, 2010). To mark content relevant to 'Stuttgart 21' both protesters and supporters used the hashtag #s21. In their use of Twitter the protests against 'Stuttgart 21' followed other political campaigns in Germany and Austria; beginning in the summer of 2009 with the #zensursula campaign against a law enabling the blocking of access to websites hosting child pornography (Bieber, 2010: 54-60), the #yeaahh flashmobs during the 2009 campaign for the federal election in Germany (Jung Herr, 2012), the #unibrennt protests for better university education (Maireder and Schwarzenegger, 2011) and the supporter campaign of Joachim Gauck (#mygauck) during the run up to the election of Germany's Bundespräsident (President) in 2010 (Hoffmann, 2010). Thus when the protesters against 'Stuttgart 21' started using social media the use of online tools for political movements was well established in Germany and applicable to the protests in Stuttgart (Schimmelpfennig, 2010; Stegers, 2010).

Social media activity reached its high point in reaction to the events during the night of September 30. Long before TV stations started to cover the protests, protesters themselves streamed video footage via mobile phones (Kuhn, 2010). An aggregation site started to collect relevant videos (<http://www.cams21.de>) to provide a ‘mosaic’ of the events (Wienand, 2010). Together with Twitter messages from the ground these videos documented the events as they unfolded and were quickly linked by German Twitter users. This led to the high visibility of the events well beyond the immediate vicinity of Stuttgart (Reißmann, 2010). The intensive coverage on blogs, Twitter, and Facebook quickly led social media users in Germany to take notice of the events in Stuttgart. An impressive amount of ad hoc analyses and link lists followed, which documented and collected the reactions online (Bunse, 2010; Pfeiffer, 2010a; Pfeiffer, 2010b).

Both, the nature of the event —a political protest that during the run of two days went through several discrete stages— and the intensive coverage of the events by social media users make the protests of the night of September 30 to October 1 an ideal topic to test the potential of event detection with Twitter data.

DATA ACQUISITION AND PREPARATION

As described above we used the Twitter streaming API to collect the data for this chapter. We focus on messages sent by Twitter users in Germany. Twitter does not require its users to state their nationality or current location reliably. The service merely encourages its users to provide some information relating to their location on their user profiles along with their local time zone. We used this information and Twitter’s random sample stream to construct our sample of German Twitter users. We approached this task in three stages:

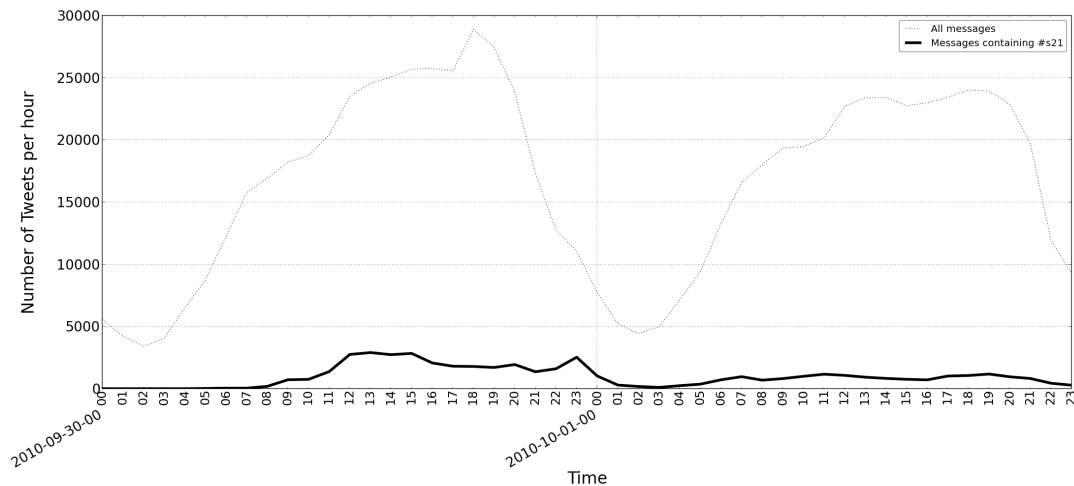
1. We collected random tweets from the Twitter streaming API;
2. We checked for hints regarding the nationality of the users posting these random tweets. In our case, we performed three checks. We assumed users to be German if they matched any of the following criteria: (a) they had their location set to one of Germany's 10.000 most populous cities, (b) their timezone contained 'Germany', 'Deutschland' or 'Berlin' or (c) they used the letter 'ß' (a letter only used in Germany and Austria).
3. Once we identified a user as German we collected their followers and those users they themselves followed. With these users we then ran the checks of step 2. This first-degree snowball sample significantly sped up the bootstrapping process and served to identify users who posted less often (and hence whose messages might not appear in Twitter's random sample).

We stopped this sampling process only after three consecutive days yielded less than a 0,1 percent increase in the number of users identified as German. Once the size of the identified German Twitter population had stabilised, we ranked the users by the number of their followers. The top 80.000 Twitter users, identified by this procedure, constitute our sample of Germany's most prominent Twitters users. For this group, we analysed all published messages on 30 September 2010 and 1 October 2010 (for a more detailed description of the sample as well as empirical tests, see Jürgens 2010).

On these two days the 80.000 German Twitter users in our sample posted 803.201 tweets. 7.793 Twitter users of our initial sample of 80.000 posted at least one message containing the hashtag #s21. In total 46.789 Twitter messages included #s21 (see Chart 1). When looking at the chart we see a strong cyclical pattern of Twitter messages corresponding with day and night rhythms, while most Twitter

messages were posted during working hours (this corresponds with findings of Golder and Macy, 2011).

Chart 1: All messages on September 30 and October 1, 2010 by the 80.000 users in our sample compared to all messages by them containing #s21



In this timeframe #s21 was by far the most popular #hashtag (46.789 mentions), followed by the longstanding Twitter usage convention ‘follow friday’ #ff (12.637 mentions), and hashtags identifying tweets commenting on astrology such as #176 (9.639 mentions), #ascendant (4.481 mentions) or #mediumcoeli (4.068 mentions). So while only roughly ten per cent of the Twitter users in our sample used Twitter to comment on the events in Stuttgart, on the days in question #s21 was the single most talked about topic on Twitter (see Chart 2).

When examining messages containing #s21 we find that a large proportion of these tweets are @messages or retweets. Of the total number of #s21 tweets (46.789) 3.389 messages were @messages while 29.138 were retweets. It is interesting to note that the number of @messages per hour remained relatively stable while the number of retweets was highly fluctuating often in connection with spikes in the overall volume of tweets containing #s21 (see Chart 3).

Chart 2: The five most popular hashtags in all messages in our sample on September 30 and October 1, 2010

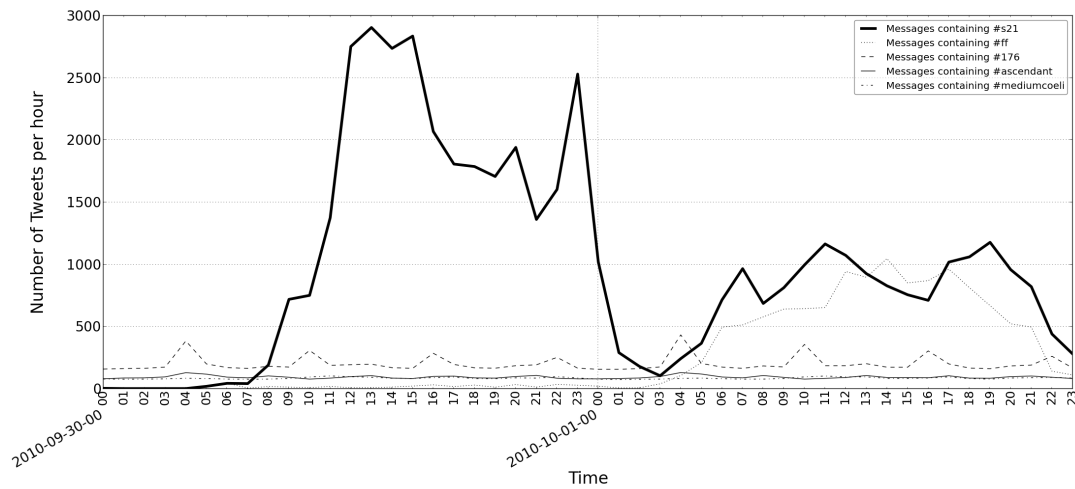
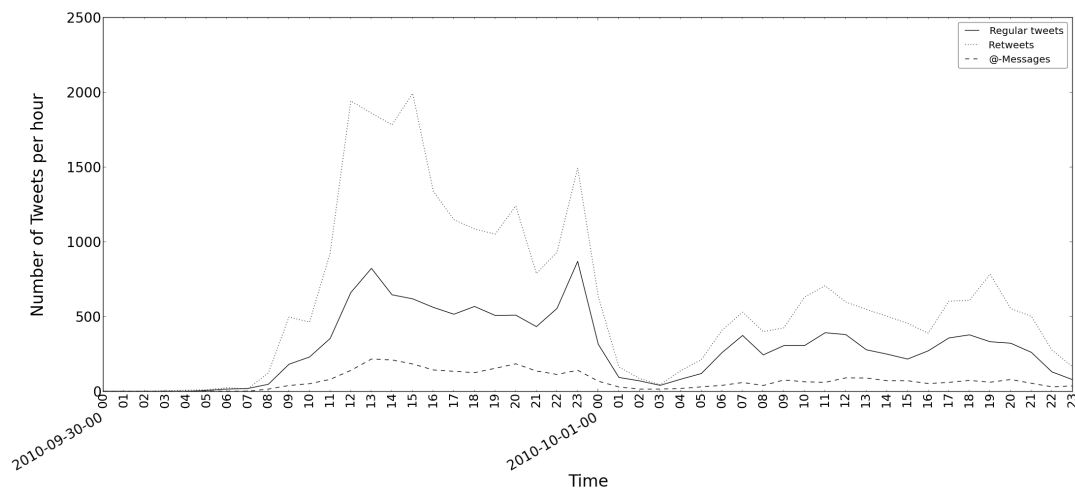


Chart 3: Comparison of normal messages, retweets and @messages containing #s21 on September 30 and October 1, 2010



STUTTGART'S BLACK THURSDAY ON TWITTER: FOUR ATTEMPTS AT EVENT DETECTION

As described above we will use four approaches for event detection with Twitter data and compare their results. These approaches are: 1. Local maxima in the volume of messages containing #s21; 2. The first occurrences of tweets that were retweeted very often during the time span of our analysis; 3. The first occurrences

of URLs that were highly linked to during the time span of our analysis; 4. Word stems with peaky characteristics.

To compare the quality of the results of these approaches we divided the two days that our analysis focused on in one-hour bins. For each of these 48 bins we calculated the word stems that were used 50 times or more in Twitter messages containing #s21. Word stems are the ‘root’ of a word that omits any inflections. For example, the stem of ‘driving’ would be ‘driv’. Collapsing words to their common root serves as a clustering methodology that maps related words to the same category. This drastically shrinks the number of possible items and enhances the power of analyses based on these data. To identify word stems we use a de-facto standard algorithm, the snowball stemmer for German that is provided by the python NLTK software package (see Bird, Loper and Klein 2009). In identifying word stems we are able to identify words describing the same context but in slightly different forms (for example ‘tree’ and ‘trees’, ‘child’ and ‘children’). The word stems used 50 times or more during each respective hour can serve as a rough indicator as to which words dominated each hour. We then used the approaches listed above to determine the time intervals of interest (see Table 1).

Table 1 shows the word stems with at least 50 mentions or more during each hour of September 30 and October 1, 2010. Already a first glance at these words suggests the nature of the events that took place during these hours. We find the German words for ‘police’, ‘park’, ‘tree’, ‘tear gas’ and ‘water cannon’. While it is difficult to discern discrete steps of the protests against ‘Stuttgart 21’ from these words alone, we are able to discern the general nature of the events. These words represent the focused attention of German Twitter users interested in the events of ‘Stuttgart 21’ on the days in question. The symbol ‘x’ in the columns ‘Volume’, ‘RTs’, and

Bin No.	Date	Time	Volume	RTs	Links	Word stems (50 mentions or more)
1	30 September 2010	0:00 - 1:00 AM				
2	30 September 2010	1:00 - 2:00 AM				
3	30 September 2010	2:00 - 3:00 AM				
4	30 September 2010	3:00 - 4:00 AM				
5	30 September 2010	4:00 - 5:00 AM				
6	30 September 2010	5:00 - 6:00 AM				
7	30 September 2010	6:00 - 7:00 AM	x			
8	30 September 2010	7:00 - 8:00 AM				
9	30 September 2010	8:00 - 9:00 AM				park, polizei
10	30 September 2010	9:00 - 10:00 AM		x	x	polizei, park, wasserwerf, baum, nil, schlagstock, demokrati, bestatigt
11	30 September 2010	10:00 - 11:00 AM				polizei, wasserwerf, schul, http://www.cams21.de/
12	30 September 2010	11:00 - 12:00 AM			x	wasserwerf, polizei, kind, fur, reizgas, traenengas, schul, #polizeigewalt, 15, friedlich
13	30 September 2010	12:00 - 1:00 PM		x	x	polizei, wasserwerf, kind, http://twitpic.com/2tbto , mal, fur, heut, wurd, reizgas, bericht
14	30 September 2010	1:00 - 2:00 PM	x		x	polizei, wasserwerf, http://twitpic.com/2tbto , bild, bitt, heut, #dpa, geh, fur, mehr
15	30 September 2010	2:00 - 3:00 PM		x	x	polizei, bitt, mehr, prot, heut, wasserwerf, stuttgart, reizgas, beim, uhr
16	30 September 2010	3:00 - 4:00 PM	x	x		heut, mensch, uhr, 1000, krankenhaus, uberlastet, prot, augenverletzt, erlitt, eil
17	30 September 2010	4:00 - 5:00 PM				heut, uhr, mahnwach, polizeigewalt, wurd, polizei, prot, viel, polizeieinsatz, 20
18	30 September 2010	5:00 - 6:00 PM		x		park, heut, kind, verletzt, uhr, 100, schadelbasisbruch, mahnwach, polizei, sanis
19	30 September 2010	6:00 - 7:00 PM		x	x	#swr, kind, demokrati, polizei, heut, geht, fur, #piraten, demo, mahnwach

20	30 September 2010	7:00 - 8:00 PM		x		heut, mehr, eigent, wurd, burg, verletzt, egal, dass, pro, #piraten
21	30 September 2010	8:00 - 9:00 PM	x	x		polizist, rech, fur, demokrat, eigent, heut, steh, gegenub, hundert, merk
22	30 September 2010	9:00 - 10:00 PM		x	x	polizei, fur, rech, heut, mal, kind, burg, eigent
23	30 September 2010	10:00 - 11:00 PM			x	polizei, polizeifunk, park, geht, fallarbeit, beginn, fur, mal, demonstrant, #cdu
24	30 September 2010	11:00 - 12:00 PM	x	x	x	baum, polizei, gasmask, kommt, polizist, erst, gefallt, rech, innenminist, baumfallfirma
25	01 October 2010	0:00 - 1:00 AM				tot, heut, mal, #cdu, fur
26	01 October 2010	1:00 - 2:00 AM				
27	01 October 2010	2:00 - 3:00 AM				
28	01 October 2010	3:00 - 4:00 AM				
29	01 October 2010	4:00 - 5:00 AM				
30	01 October 2010	5:00 - 6:00 AM				
31	01 October 2010	6:00 - 7:00 AM				heut
32	01 October 2010	7:00 - 8:00 AM	x			bundestag, debatt, #bundestag, heut, grun, fur, ab, antrag, schwarzgelb, uber
33	01 October 2010	8:00 - 9:00 AM				ab, heut, polit
34	01 October 2010	9:00 - 10:00 AM			x	heut, uhr, fur
35	01 October 2010	10:00 - 11:00 AM		x	x	rucktritt, fur, ford, heut, innenminist, #rech, rech, bahnhof, polizeieinsatz, brutal
36	01 October 2010	11:00 - 12:00 AM	x	x	x	fur, bitt, http://youtu.be/W1UYd5LDQXA , uhr, uber, video, weiterverbreit, geht, gewalt, phoenix, video
37	01 October 2010	12:00 - 1:00 PM				fur, heut, bitt, uhr, geht, http://youtu.be/W1UYd5LDQXA , video
38	01 October 2010	1:00 - 2:00 PM				fur, heut, geht, bos, schlagzeil, printmedi, preis, #ftd
39	01 October 2010	2:00 - 3:00 PM				fur, heut, geht, bitt, polizei, schlagzeil, preis

40	01 October 2010	3:00 - 4:00 PM			heut, fur, bitt, uber, abstimm
41	01 October 2010	4:00 - 5:00 PM	x		demo, heut, fur, 1900, schlossgart, livestream, ab, www.polizei.co, www.krieg.co, http://bit.ly/a0aFpA
42	01 October 2010	5:00 - 6:00 PM			heut, polizei, gest, schon, #rech, #mappus, fur, wurd, baum
43	01 October 2010	6:00 - 7:00 PM		x	eisenbahnbundesamt, demo, frau, seit, fur, heut, 100000, gest, fall, db
44	01 October 2010	7:00 - 8:00 PM	x	x	mappus, polizei, 100000, demonstrant, ja, uhr, uber, rucktritt, konnt, gesprach
45	01 October 2010	8:00 - 9:00 PM			mappus, slomka, #zdf, #mappus, wurd, frau, heutejournal, fur, mal, gesprach
46	01 October 2010	9:00 - 10:00 PM			mappus, fur, neu, lug, polizei, masseein, abstand, #mappus, schon
47	01 October 2010	10:00 - 11:00 PM			fur
48	01 October 2010	11:00 - 12:00 PM			

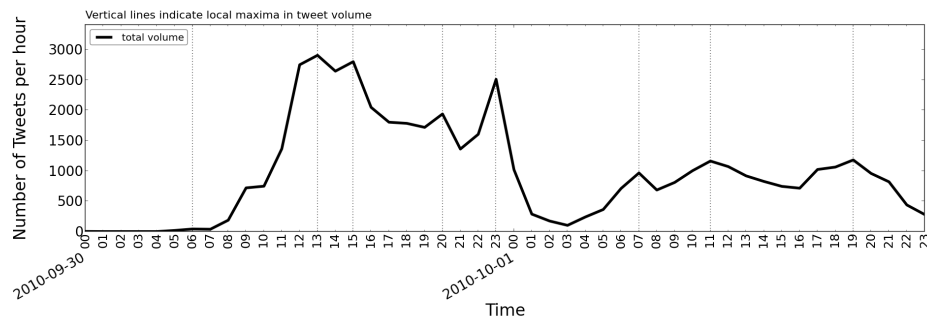
Table 1: Word stems used at least 50 times in any given hour

‘Links’ indicate those hours that were identified as significant by the respective event detection approach used. The following section develops each of these approaches and the results of their application in our data.

Message Volume

One simple approach is to examine the fluctuations in the volume of messages containing #s21. In Chart 4 we graph the total volume of messages containing #s21 in any given hour of the two days in question. The lines mark local maxima were the volume reached relative peaks.

Chart 4: Tweet volume per hour, the vertical lines identify significant hours as detected by local maxima, local maxima are detected if tweet volume is lower both before and after a local point



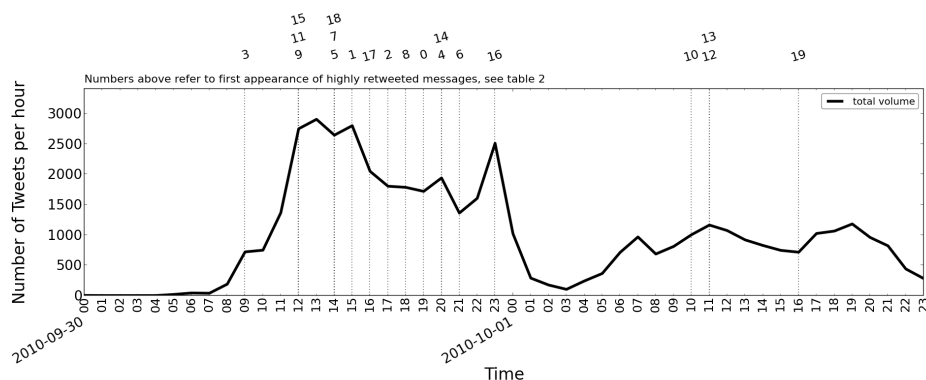
The chart shows that the volume of Twitter messages containing the hashtag #s21 follows a clear day and night rhythm. Still, especially on September 30 the overall volume is much higher than the following day and it shows distinctive patterns between 12:00 and 3:00 PM and again at 11:00 PM. These peaks are clearly detected by an analysis based on local maxima. In Table 1 we showed the one-hour bins in which the local maxima fall by the value 'x' in the column 'Volume'. When looking at the corresponding word stems that were used 50 times or more in these bins we find that local maxima correspond with distinct events during the protest. For example the peak around 1:00 PM corresponds with heavy clashes between police and protesters, among them school children. Messages dealing with this event dominate the Twitter discourse over the following hours. The sudden peak at 11:00 PM the same day corresponds with a new development in the protest. At that time construction workers started to cut down trees under police protection. This was accompanied by heavy protests and clashes between police and protesters. Thus it can be said that the analysis of local maxima in the volume of Twitter messages

commenting on a given topic can provide a first understanding of the development of the event in question. They indicate important timespans during the event while the words used most often during these timespans allow a view on the elements of the event that held public attention.

Retweets

Another approach to the detection of events is the identification of messages that were retweeted often. We identified the 20 Tweets in our data set that were retweeted most often during the two days in question (see Table 2). The number we chose, 20, is rather arbitrary. Still, we find that in our case the number of retweets a message received stabilised after the twentieth rank. For other analyses or other time spans a different number of retweets might be better suited. The frequency of retweets ranged from 271 at the top to 75 at the bottom of our 20 most often retweeted messages. We then identified the time these Tweets were originally posted and checked what kind of information these tweets held on the protests (see Chart 5).

Chart 5: Tweet volume, the vertical lines identify significant hours as detected by the first appearance of the 20 most retweeted messages



ID	Count	Date	Time	Tweet	Type
0	271	30 September 2010	7:46:09 PM	RT @dingler_g4: Pro oder contra #s21 ist eigentlich egal. Man prügelt seine Bürger nicht. Punkt.	Commentary
1	248	30 September 2010	3:08:30 PM	RT @tazgezschwitscher: EIL (dapd) +++ 1.000 Menschen haben Augenverletzungen erlitten, Krankenhäuser in Stuttgart überlastet #s21	Misinformation
2	150	30 September 2010	5:50:30 PM	RT @triffy: Stuttgart 21: Kirche zeigt sich empört, das Verprügeln kleiner Jungs sei ihre Aufgabe! #s21, 150, ,	Satire
3	148	30 September 2010	9:08:20 AM	RT @fasel: Schlagstöcke und Wasserwerfer, die Grundpfeiler einer gesunden Demokratie #S21	Commentary
4	139	30 September 2010	8:10:46 PM	RT @ChrMll: Wie ich mir 11880 merke? Hundert 11-jährige Demokraten stehen 88 Polizisten gegenüber und haben 0 Chance. #s21	Satire
5	127	30 September 2010	2:53:45 PM	RT @zebramaedchen: Wenn Iraner jetzt aus Solidarität unter Location "Stuttgart" eintragen, wird es ernst. #s21	Satire
6	124	30 September 2010	9:23:38 PM	RT @Schmidtlepp: Bahnhof des himmlischen Friedens = 天安火车站 #S21	Satire
7	123	30 September 2010	2:26:35 PM	RT @C_Holler: BREAKING+++ Demokratiefeindliche Ökostalinisten behindern Stuttgarts Polizei beim Blumengießen. #S21	Satire
8	121	30 September 2010	6:24:17 PM	RT @eldersign: In einer Demokratie kann man bedenkenlos Kinder mit auf eine Demo nehmen, in einem #Polizeistaat nicht. #S21	Commentary
9	121	30 September 2010	12:55:46 PM	RT @forschungstorte: dieses bild soll um die welt gehen: #s21 http://twitpic.com/2tbtod #dpa - sowas passiert in #deutschland	Media
10	113	01 October 2010	10:52:18 AM	RT @saschalobo: Das Gegenteil von höflich heisst seit gestern bahnhöflich. #S21	Satire
11	103	30 September 2010	12:24:11 PM	RT @phlox81: Und vergesst nicht, wenn es nach der #CDU ginge, würde da jetzt auch die #Bundeswehr mitmischen! #s21	Satire
12	98	01 October 2010	11:03:59 AM	RT @NineBerry: Bitte dieses Video weiterverbreiten: http://youtu.be/W1UYd5LDQXA #s21	Protestmedia
13	94	01 October 2010	11:19:46 AM	RT @pillenknick: Der Preis für die böseste Schlagzeile zu #S21 in Printmedien geht an die #FTD: http://twitpic.com/2tlg9t	Media
14	94	30 September 2010	8:19:11 PM	RT @ChrMll: Wie praktisch: Wenn die Revolution am Sonntag kommt, bleibt sogar der Nationalfeiertag der gleiche. #s21	Satire
15	91	30 September 2010	12:46:44 PM	RT @ZDFonline: Der Konflikt um Stuttgart 21 eskaliert - laut Augenzeugen mit Pfefferspray und Wasserwerfern http://bit.ly/czGpM7 #s21 #p ...	Media
16	90	30 September 2010	11:27:12 PM	RT @tauss: Die #s21 Baumfällfirma kommt übrigens aus dem Wahlkreis von Innenminister Heribert Rech (Karlsdorf b. Bruchsal)	Information
17	87	30 September 2010	4:59:14 PM	RT @tazgezschwitscher: 100 verletzte Kinder, 1 Schädelbasisbruch – und die Sanis dürfen nicht in den Park. (Quelle: Parkschützer) #s21 http ...	Misinformation
18	82	30 September 2010	2:13:06 PM	RT @abrisaufstand: bitte! wir brauchen mehr Sanitäter & Ärzte #S21 sie sprühen Reizgas willkürlich und massiv in die Menge #S21	Information
19	75	01 October 2010	4:52:52 PM	RT @KnirpsStore: Schirm durch einen Wasserwerfer kaputt gegangen? Wir unterstützen gerne friedliche Demos. Mail an s21@knirps-store.de < mailto:s21@knirps-store.de > #s21	Satire

Table 2: Top 20 RTs in Twitter messages on September 30 and October 1, 2010 containing the #keyword #s21

As shown in the columns ‘Volume’ and ‘RTs’ of Table 1 the 20 most popular retweets identified more time bins than the previous approach through local maxima. Still, there is no large difference in the time bins identified by both methods. Messages that were highly retweeted were posted mostly during the active hours of the protest in Stuttgart. When looking at the word stems that were used in the hours indicated by popular tweets we roughly get the same picture of the protest as when looking at the word stems in hours indicated by local maxima. Also when looking at the content of the most often retweeted tweets themselves we do not get a more detailed look at the events in Stuttgart. This is largely due to the fact that only a small minority of those tweets contained actual information on the event itself. Most messages that were retweeted intensively contained either generic commentary on the events (for example ‘Pro or against #s21 does not really matter. You don’t beat up your citizens. Full stop’ (ID=0)) or satirical content (for example ‘How do I remember the 11880 [the number of Germany’s telephone information]? Hundred 11-year old democrats stand against 88 policemen and have 0 chance. #s21’ (ID=4)). While this content is highly popular among retweeters, maybe because it crystallises their thinking or reaction to the events, these tweets do not help researchers interested in the development of the unfolding protest. However, when looking at tweets that were less often retweeted we increasingly find tweets of activists that contain actual information on different stages of the protest. For example, ‘The police does not let journalists through, even those with official credentials’, or ‘They said via microphone that one man has lost his eyesight through a water cannon’. So activists used Twitter consciously to distribute information about the protests and to organise it —as has also been shown in other

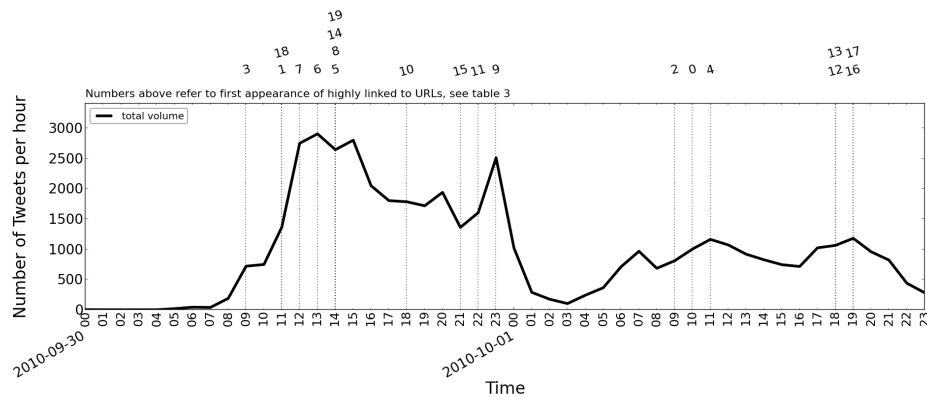
cases (Jungherr, 2009)— but these tweets did not receive as much retweet attention as tweets containing generic or satirical commentary.

Another indicator shows potential problems with expecting tweets containing actual information on the protests to be highly retweeted when examining which applications people used when tweeting about the protests, we found that applications for mobile phones, which can be used to post messages on Twitter on the go, were used to publish only about a quarter of the total tweets containing the hashtag #s21. This means that most of the volume of tweets commenting on the events actually comes from users sitting in front of desktop computers or notebooks who followed the protests from afar. For those users satirical or witty tweets are of course more attractive to retweet than messages containing procedural information. So while the time of the initial posting of popular retweets often corresponds with time bins identified by spikes in total volume of messages the actual content of the most popular retweets does not provide a better understanding of the unfolding events. While local maxima helped us understand which words were used by most users in their attempt to comment on #21, popular retweets help us understand which messages these users considered to be most representative for their own reactions to the events.

URLs

As a third approach to the identification of relevant time bins, we can use domains that were popularly linked to in tweets containing #21. After identifying the domains linked to most often in #s21 tweets we measured the time when they were first posted and marked the corresponding time bin (see Chart 6).

Chart 6: Tweet volume, the vertical lines identify significant hours as detected by the first appearance of the 20 most linked to URLs



Again we find that most time bins identified by this method correspond with those identified by local maxima or retweets (see column ‘URLs’ in Table 1). Again we find that the word stems used most often in those time bins do not provide us with very specific information about the development of the protest. So, do the linked domains tell us something about the protests that we did not know before?

In Table 3 we documented the 20 most popular domains linked to in #s21 tweets. As before, 20 is an arbitrary number that is useful in this specific analysis but that could be expanded. It is interesting to note that most linked domains do provide access to media provided by the protesters themselves. Links to video or audio live streams that document the unfolding protests are particularly popular. In this the ‘Stuttgart 21’ protests mirror other recent practices among political activists (Pickard, 2006; David, 2010). It is also interesting to note that although many of the links go to videostreams, nearly no digital photographs taken by protesters themselves are detected among the most popular links. There are some links to pictures, but these tend to be snapshots of newspaper articles or links to photographs taken by professionals. In documenting the #s21 protests on September

ID	Website	Description	Type	Link Count
0	http://www.youtube.com/verify_age?next_url=/watch%3Fv%3DW1UYd5LDQXA%26feature%3Dyoutu.be	Videoclip of protest	Protest Media	311
1	http://fluegel.tv/	Videostream of protest	Protest Media	308
2	http://www.campact.de/bahn/ml4/mailer	Campact mail campaign in reaction to protests	Campaign	246
3	http://www.cams21.de/	Videostream of protest	Protest Media	204
4	http://twitpic.com/2tlg9t	Picture of newspaper article on protest	News	203
5	http://piratenpad.de/s21	Public pad of Germany's Pirate party, used to coordinate protests	Party Content	175
6	http://bambuser.com/channel/terminal.21/broadcast/105357	Videostream of protest	Protest Media	160
7	http://www.heute.de/ZDFheute/inhalt/30/0,3672,8116958,00.html	Newscoverage of protest	News	149
8	http://www.amnestypolizei.de/	Campaign by Amnesty International for transparency of police actions	Campaign	120
9	http://fxneumann.de/2010/10/01/ohnmacht-wut-und-repraesentative-demokratie/	Blogpost in reaction to protest	Blog	103
10	http://images.zeit.de/politik/deutschland/2010-09/bg-stuttgart21-bilder/20754237-540x304.jpg	News photo of protest	News	101
11	http://www.ustream.tv/channel/kaputtgart-21	Videostream of protest	Protest Media	96
12	http://twitpic.com/2tp8ui	Twitpic of official document concerning the protests	Protest Media	95
13	http://www.piratenpartei.de/Pressemitteilung-100930-PIRATEN-entsetzt-ueber-Traenengaseinsatz-gegen-Schueler-bei-S21-Demo	Official press statement of Germany's Pirate party in reaction to the police action	Party Content	85
14	http://twibbon.com/join/Oben-bleiben-s21	Support Twibbon for protesters against #s21	Protest Media	83
15	http://twitpic.com/2tpka5	Twitpic of protest	Protest Media	81
16	http://twitpic.com/2tpk9o	Twitpic of protest	Protest Media	80
17	http://twitpic.com/2tb48b	Twitpic of protest	Protest Media	72
18	http://taz.de/!59135/	Newscoverage of protest	News	70
19	http://wiki.piratenpartei.de/Landesverband_Baden-W%C3%BCrttemberg/Arbeitsgruppen/Presse/S21	Wiki of Germany's Pirate party collecting press reactions to Stuttgart21	Party Content	68

Table 3: Top 20 websites linked in Twitter messages on September 30 and October 1, 2010 containing the #keyword #s21

30 and October 1, 2010 digital photography by activists themselves had only a marginal role. Although the linked domains offered other Twitter users on that evening a more detailed look at the unfolding events they do not provide much information for researchers looking for the development of the process. Still, popularly linked domains provide us with an impression of what Twitter users were paying attention to while publicly commenting on the events.

With regard to the questions stated above we find that popular URLs provide us with similar time bins as other event detection approaches and that content of established media are not overrepresented. Social media data thus seems to offer a view of unfolding events independent of media accounts.

Peakiness

In our final approach we identify relevant time bins by the relative frequency of word stems. Following an approach initially proposed by Shamma, Kennedy and Churchill (2011) we identified those word stems that were used at least 100 times during the two days and received at least 50 percent of their mentions during one single hour (see Table 4). Shamma et al. call this pattern ‘peaky’.

We used the hours indicated by the appearance of word stems with peaky characteristics and compared them to the hours indicated by the three previous approaches (see Table 1). Again we find that most hours identified by this approach correspond quite well with those identified by the other approaches (see Chart 7). Since the calculation of peakiness occurs in slices of one hour each, the indicated moments of relevant activity have a maximum precision of one hour.

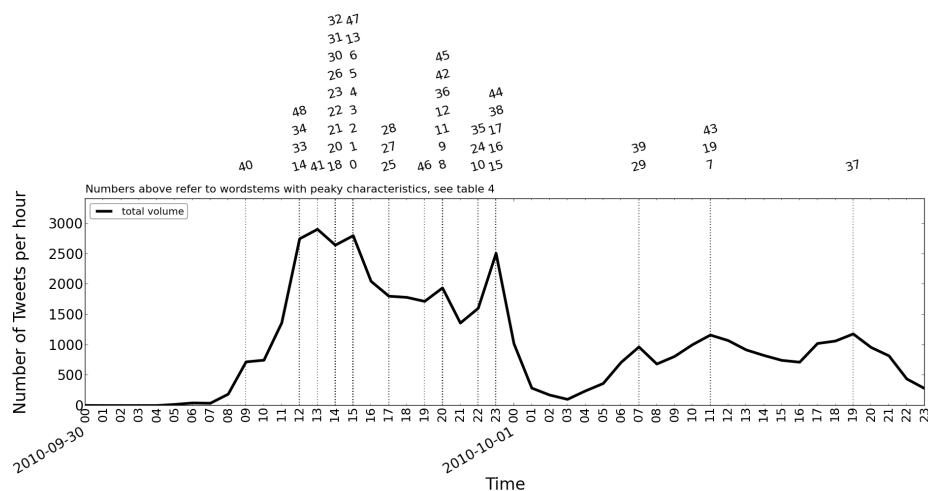
ID	Word stem	Date	Time	Example	Type
0	krankenhaus	30 September 2010	3:00 - 4:00 PM	RT @Tschuly82: Ungesicherte Infos ca. 1000 Verletzte #s21 in den Krankenhäusern in #Stuttgart	Misinformation
1	überlastet	30 September 2010	3:00 - 4:00 PM	bitte NICHT mit Reizgasverletzungen in die Stuttgarter Krankenhäuser; diese sind komplett überlastet #S21 bitte RT	Misinformation
2	1000	30 September 2010	3:00 - 4:00 PM	meinen 1000ten tweet widme ich dem widerstand in stuttgart: nicht aufgeben, weitermachen! #S21	Non-specific
3	augenverletzt	30 September 2010	3:00 - 4:00 PM	Augenverletzungen dokumentieren: Augenarztpraxis am Olgaek, Charlottenstraße 23 #S21	Information
4	eil	30 September 2010	3:00 - 4:00 PM	RT @Roland_Veile: #S21: Bitte sofort alle in den #Park	Non-specific
5	erlitt	30 September 2010	3:00 - 4:00 PM	EIL (dapd) +++ 1.000 Menschen haben Augenverletzungen erlitten, Krankenhäuser in Stuttgart überlastet #s21	Misinformation
6	dapd	30 September 2010	3:00 - 4:00 PM	EIL (dapd) +++ 1.000 Menschen haben Augenverletzungen erlitten, Krankenhäuser in Stuttgart überlastet #s21	Misinformation
7	http://youtu.be/W1UYd5LDQXA	01 October 2010	11:00 - 12:00 AM	Bitte dieses Video weiterverbreiten: http://youtu.be/W1UYd5LDQXA #s21	Protest Media
8	chanc	30 September 2010	8:00 - 9:00 PM	RT @stoddnet: Keine Chance, wenn selbst Kinder geschlagen werden #s21	Non-specific
9	88	30 September 2010	8:00 - 9:00 PM	@Debe1887 Und wieder wirst du bestätigt in einigen deiner Aussagen. #S21	Non-specific
10	polizeifunk	30 September 2010	10:00 - 11:00 PM	Darf man eigentlich einen Livestream des Stuttgarter Polizeifunks vertwitern? http://www.ustream.tv/recorded/9913851 #s21	Protest Media
11	11jährig	30 September 2010	8:00 - 9:00 PM	"Tagesspiegel" berichtet von Gewalt gegen 11jährige Schüler und 60 jährige Frauen. http://bit.ly/a9d97W #s21	Media
12	11880	30 September 2010	8:00 - 9:00 PM	Wie ich mir 11880 merke? Hundert 11-jährige Demokraten stehen 88 Polizisten gegenüber und haben 0 Chance. #s21	Non-specific
13	http://twitpic.com/2tcutw	30 September 2010	3:00 - 4:00 PM	Hehe: Wer nicht hört bekommt auf die Fresse! RT @scheelm: http://twitpic.com/2tcutw #s21	Media
14	hinlang	30 September 2010	12:00 - 1:00 PM	#s21 proteste in #stuttgart: ein beamtensprecher erklärt auf #taz online, die polizei kann ruhig mal hinlangen: http://tinyurl.com/38rzazd	Media
15	baumfallfirma	30 September 2010	11:00 - 12:00 PM	Baumfällfirma aus Ba-Wü: Gredler & Söhne, Waldstrasse 17, 76689 Karlsdorf-Neuthard, Tel: 07251-9443-0 Fax: -9443-22, #S21	Information
16	wahlkreis	30 September 2010	11:00 - 12:00 PM	RT @lutz__h: Wahlkreisbüro von Stefan #Mappus, #CDU, Pforzheim: 07231 / 1458-0 #S21	Information
17	gasmask	30 September 2010	11:00 - 12:00 PM	gasmasken - Polizei hat Gasmasken #s21	Information
18	behind	30 September 2010	2:00 - 3:00 PM	schlagzeile morgen in der BILD: "wildgewordene schüler und rentner behindern friedliche baumfällarbeiten" #s21	Satire
19	weiterverbreit	01 October 2010	11:00 - 12:00 AM	RT @JulianMuetsch: Mahnwache zu #S21 um 19Uhr vor dem Mannheimer HBF! Bitte Weiterverbreiten!	Information
20	breaking	30 September 2010	2:00 - 3:00 PM	Breakingnews: In #Stuttgart gehen derzeit #Polizisten brutal gegen friedliche Demonstranten - darunter viele #Schüler - vor. #S21	Non-specific
21	demokratiefeind	30 September 2010	2:00 - 3:00 PM	RT @C_Holler: BREAKING+++ Demokratiefeindliche Ökotalinisten behindern Stuttgarts Polizei beim Blumengießen. #S21	Non-specific

22	blumengiess	30 September 2010	2:00 - 3:00 PM	RT @C_Holler: BREAKING+++ Demokratiefeindliche Ökotalinisten behindern Stuttgarts Polizei beim Blumengießen. #S21	Satire
23	okotalinist	30 September 2010	2:00 - 3:00 PM	RT @C_Holler: BREAKING+++ Demokratiefeindliche Ökotalinisten behindern Stuttgarts Polizei beim Blumengießen. #S21	Satire
24	beginn	30 September 2010	10:00 - 11:00 PM	die parkräumung und baumfällung beginnt. #S21	Information
25	schadelbasisbruch	30 September 2010	5:00 - 6:00 PM	RT @hellertaler: fluegel.tv: "Über 100 verletzte Kinder, ein Schädelbasisbruch, diverse sonstige heute ist was zerbrochen. Vertr.. #s21	Misinformation
26	arzt	30 September 2010	2:00 - 3:00 PM	@Kyra2001: #S21 Augenärzte und Ärzte werden noch gebraucht. Vor allem #Augenärzte Vor Ort! Augenverletzungen S21	Information
27	parkschutz	30 September 2010	5:00 - 6:00 PM	RT @Cymaphore: Kranwagen im Schlosspark in #Stuttgart wird blockiert #S21 #K21 #Parkschützer	Information
28	sanis	30 September 2010	5:00 - 6:00 PM	Verletzte: Demosanis: Biergarten, Rotes Kreuz: Cannstatter Ende des mitt. Schlossgartens, Rettungswagen am Südausgang #S21 bitte RT	Information
29	#bundestag	01 October 2010	7:00 - 8:00 AM	#S21: Bitte alle #Berliner und #Berlinerinnen. Demonstriert bitte vor dem #Bundestag gegen Stuttgart21.	Information
30	willkur	30 September 2010	2:00 - 3:00 PM	bitte! wir brauchen mehr Sanitäter & Ärzte #S21 sie sprühen Reizgas willkürlich und massiv in die Menge #S21	Information
31	sanitat	30 September 2010	2:00 - 3:00 PM	Das Zelt der Sanitäter ist auf der Wiese zwischen Biergarten und Cafe Nil #S21	Information
32	spruh	30 September 2010	2:00 - 3:00 PM	jetzt prügeln und sprühen sie auch noch aufs Deeskalationsteam ein #s21 #WTF	Information
33	#bundeswehr	30 September 2010	12:00 - 1:00 PM	Und vergesst nicht, wenn es nach der #CDU ginge, würde da jetzt auch die #Bundeswehr mitmischen! #s21	Commentary
34	mitmisch	30 September 2010	12:00 - 1:00 PM	RT @phlox81: Und vergesst nicht, wenn es nach der #CDU ginge, würde da jetzt auch die #Bundeswehr mitmischen! #s21	Commentary
35	fallarbeit	30 September 2010	10:00 - 11:00 PM	fällarbeiten - An alle Stuttgarter Parkschützer: AUF IN DEN PARK!!! Baumfällarbeiten beginnen nun! http://www.parkschuetzer.de/webcam #s21 #k21	Information
36	sonntag	30 September 2010	8:00 - 9:00 PM	Heute die Demokratie mit Füßen treten und am Sonntag Geschwollene Reden über Einheit in Frieden und Freiheit halten ... #s21	Non-specific
37	mind	01 October 2010	7:00 - 8:00 PM	RT @Sugg__: Mindestens 2 Wasserwerfer sind schon im Park!!! #S21 #Aufstand #WehrtEuch	Non-specific
38	karlsdorf	30 September 2010	11:00 - 12:00 PM	Baumfällfirma aus Ba-Wü: Gredler & Söhne, Waldstrasse 17, 76689 Karlsdorf-Neuthard, Tel: 07251-9443-0 Fax: -9443-22, #S21	Information
39	antrag	01 October 2010	7:00 - 8:00 AM	Begründung war übrigens, das sei eine Spontandemo, die müsste schriftlich beantragt werden. #dortmund #s21	Information
40	nil	30 September 2010	9:00 - 10:00 AM	RT @robin_wood: #Wasserwerfer am Café Nil im Park. #S21	Information
41	seltsam	30 September 2010	1:00 - 2:00 PM	Gute Nacht, seltsame Welt! Denk' ich an Deutschland in der Nacht, so bin ich um den Schlaf gebracht... #Polizeistaat #S21	Non-specific
42	praktisch	30 September 2010	8:00 - 9:00 PM	@JoGoebel naja, in einer repräsentativen demokratie äußern sich praktisch alle parteien stellvertretend für gewisse gruppen, oder? #s21	Non-specific
43	phoenix	01 October 2010	11:00 - 12:00 AM	Stuttgart #s21 live im Bundestag! #phoenix	Information
44	bruchsal	30 September 2010	11:00 - 12:00 PM	bruchsaler - Schichtwechsel hinterm Zaun. Rheinland-Pfalz wurde durch aggressive Bruchsaler abgelöst. #S21	Information

45	nationalfeiertag	30 September 2010	8:00 - 9:00 PM	Wie praktisch: Wenn die Revolution am Sonntag kommt, bleibt sogar der Nationalfeiertag der gleiche. #s21	Non-specific
46	sekundentakt	30 September 2010	7:00 - 8:00 PM	Im sekundentakt kommen Tweets. Ab in die Trending Topic, zeigt der Welt was hier passiert! #s21	Non-specific
47	sek	30 September 2010	3:00 - 4:00 PM	RT @HMSzymek: Die "Wichtigen" in Zivil sind eher SEKs bei der Menge. #s21	Information
48	http://bit.ly/czGpM7	30 September 2010	12:00 - 1:00 PM	Der Konflikt um Stuttgart 21 eskaliert - laut Augenzeugen mit Pfefferspray und Wasserwerfern http://bit.ly/czGpM7 #s21 #parkschützer	Media

Table 4: Peak Word stems, hashtags, and links that appear at least 100 times in the text corpus and who are used more than 50% in one specific hour

Chart 7: Tweet volume, the vertical lines identify significant hours as detected by word stems with peaky characteristics, peakiness is assumed at 0.5 (50% of occurrences within this hour) for all word stems that appeared at least 100 times on September 30 and October 1, 2010



Following the pattern of peaky word stems, four major phases of the protests can be identified. After a first onset around 9 AM which is connected to a first sighting of water cannons, the main phase of the demonstration starts around 12 AM. It is characterised by content that comments rather than reports (for example ‘breaking’, entry 20 in Table 4). The main theme —apart from the general mentioning of protests— are reported injuries through tear gas and water cannons (‘sanitat’ = medic, entry 31 in Table 4). From 3 PM onwards, links to live streams start to appear (items 0-6 in chart 7). The intermediary period (5-7 PM) is dominated by

humorous meta-coverage. The second major sub-event happens during the night, when trees are first being chopped down (‘fallarbeit’ = logging work, item 35 in table 4). The following day is characterised by meta-coverage and links to media stories (for example ‘phoenix’, a German TV station, item 43 in table 4).

Overall, isolating peaky word stems provides researchers with a general yet relevant summary of the event in question. In most cases however, the context from entire tweets is still needed in order to interpret the meaning of word stems.

CONCLUSION: EVENT DETECTION WITH SOCIAL MEDIA DATA

We presented four distinct approaches for the detection of events in social media data. All four approaches —based on volume alone, salient retweets, salient URLs and a metric called ‘peakiness’ (Shamma, Kennedy and Churchill 2011)— were applied to the same two-day dataset documenting protests reacting to the construction of a new train station in Stuttgart, Germany.

Even at first glance, it is apparent that tweets addressing the protests (containing the #s21 hashtag) seem to develop parallel to the actual events. During crucial phases of the protest the volume of messages commenting on #s21 rises. This extends a basic observation: at its most simple level, the volume of Twitter messages over time mirrors basic human activity patterns—high volume during waking hours and on weekends, low volume during the night. What do the four approaches chosen by us add to that basic observation?

1. Volume: the analysis of local maxima in the volume of tweets succeeds in finding significant time bins. Times indicated by this approach coincide with the most active phases of the protests. It should be noted that while in this case activity

online corresponds with activity offline this might not always be the case. That is, in general protests might occur but might go unnoticed on Twitter. Also Twitter users might discuss protests without any corresponding offline event.

2. Retweets: examining the tweets that during the run of our analysis were retweeted the most we also find a temporal structure of the protests. We find that the time bins marked by the date of the original publication of these tweets largely match the time bins indicated by local maxima in message volume. Beyond the identification of key points in the time line, retweets also supply a first glance at what is happening at these moments. Still, a cursory analysis of the 20 most retweeted messages (cf. table 2) shows that most of the time their content does not address the various stages of the protest but instead mostly offers humorous commentary. Thus the content of the most salient tweets does not offer a map of the protests but instead provides an overview of the social media objects (for example pictures, links, jokes) connected to the protests in Stuttgart that were amplified by commentators of the event and not the view of activists present at the protests.

3. URLs: examining the times when the most salient URLs were initially linked to offers another temporal structure of the events. The results for the Stuttgart dataset show that Twitter messages did not trail accounts of traditional media. The time bins indicated by popular URLs overlap often with the time bin indicated by local maxima in message volume. Many of the referenced websites provide live coverage of the protests through video streams (cf. Table3, column ‘type’ and there all entries that were labelled ‘protest media’). So we can clearly state that at least some users used Twitter in combination with other social media channels to cover the events as they were unfolding. While we still cannot claim that the URL analysis yields an

appropriate overall picture of the event itself, it has become clear that (at least in this case) Twitter messages tracked the protest dynamics with little to no delay.

4. Peakiness: so far it has become clear that the approaches described above merely offer coarse maps of the actual event. The peakiness approach offers a somewhat more detailed account of the events. The approach clearly identified time bins that corresponded with important stages in the development of the protest. In addition, it identifies and locates many peaky word stems that are descriptive of important changes in the situation. Good examples for this are the mentioning of gasmasks, the moment when trees were first chopped down ('Fällarbeiten') as well as reported eye injuries ('Augenverletzung') (cf. Table 4). In contrast to the approaches focusing on salient tweets and URLs—that rely on the prominent visibility of a small subset of all messages (by an even smaller subset of users)—peakiness is able to detect meaningful trends based on messages posted by a widely distributed group of users. Therein lays its biggest advantage: it is not constrained to the recognition of the few most prominent phenomena. Even if thousands of people decided to independently tweet about a certain moment without referencing each other, the peak would still be detected—as long as they use roughly the same vocabulary. A caveat remains in that many of the peaky word stems fail to provide meaningful information on their own. As such, the method is not suited for fully automated analyses. As one tool among many for exploratory analysis, however, it offers a different and valuable new approach to the data.

In this chapter, we investigated the analytical potential of four approaches to event detection with data collected on a social media channel, namely Twitter. We have argued that although these approaches provide a somewhat successful overview over the events, their scientific value remains limited by the lack of

representativeness of the original data set. Thorough inferences about real world protests still require direct, unbiased observations that are unavailable in social media and other media channels (confer Lang and Lang, 1953). We see the potential of analyses based on social media data mainly in their capacity to structure large quantities of unknown data and aid researchers in exploratory sighting, that — depending on the respective research question— potentially need corroboration by studies that are based on representative samples of the population under study.

REFERENCES

- Allan, J. (ed.) (2002) *Topic detection and tracking: Event-based information organization*, Boston: Kluwer Academic Publishers.
- Asur, S., and Huberman, B. A. (2010) *Predicting the future with social media*.
[Online], Available: <http://arxiv.org/abs/1003.5699>.
- Becker, H., Naaman, M., and Gravano, L. (2011) ‘Beyond trending topics: real-world event identification on Twitter’, in *Proceedings of the fifth international AAAI conference on weblogs and social media*, 438-441, Menlo Park, California: The AAAI Press.
- Bieber, C. (2010) *politik digital: Online zum Wähler*, Salzhemmendorf: Blumenkamp Verlag.
- Bilger, C., and Raidt, E. (2011) ‘Schwarzer Donnerstag: Ein lauter Tag hallt nach, Stuttgarter Zeitung’, [Online], Available: <http://www.stuttgarter-zeitung.de/inhalt.schwarzer-donnerstag-ein-lauter-tag-hallt-nach.94770415-a4dd-4957-8422-a689eaa2909e.html>.
- Bird, S., Loper, E. and Klein, E. (2009) *Natural language processing with Python*. Sebastopol, CA: O’Reilly Media.

- Brandt, P. T., Freeman, J. R. and Schrodtt, P. A. (2011) 'Real time, time series forecasting of inter- and intra-state political conflict', *Conflict Management and Peace Science* 28, 41-64.
- Bruns, A. and Burgess, J. (2011) '#Ausvotes: How Twitter covered the 2010 Australian federal election', *Communication, Politics & Culture* 44(2), 37-56.
- Bruns, A. and Liang, Y. E. (2012) 'Tools and methods for capturing Twitter data during natural disasters', *First Monday* 17(4), [Online], Available: <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3937/3193>.
- Bunse. V. (2010) '#S21: Man prügelt seine Bürger nicht', ... *Kaffee bei mir?* [Online], Available: <http://opalkatze.wordpress.com/2010/09/30/s21-man-prugelt-keine-burger/>.
- Busemann, K. and Gscheidle, C. (2012) 'Web 2.0: Habitualisierung der Social Communitys', *Media Perspektiven* 7-8, 380-390.
- Cha, M., Haddadi, H., Benevenuto, F. and Gummadi, K. P. (2010) 'Measuring user influence in Twitter: The million follower fallacy', in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 10-17, Menlo Park, California: The AAAI Press.
- Chakrabarti, D. and Punera, K. (2011) 'Event summarization using tweets', in *Proceedings of the fifth international AAAI conference on weblogs and social media*, 66-73, Menlo Park, California: The AAAI Press.
- Chew C. and Eysenbach G. (2010) 'Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak', *PLoS ONE* 5/11, e14118.

- Crawford, K. (2009) 'Following you: Disciplines of listening in social media', *Continuum: Journal of Media & Cultural Studies* 23(4), 525-535.
- David, G. (2010) 'Camera phone images, videos and live streaming: A contemporary visual trend', *Visual Studies* 25(1), 89-98.
- Gabriel, O.W., Schoen, H. and Faden-Kuhne, K. (2013) *Die Volksabstimmung über 'Stuttgart 21'*, Leverkusen: Opladen Budrich.
- Gayo-Avello, D., Metaxas, P.T. and Mustafaraj, E. (2011) 'Limits of electoral predictions using Twitter', in *Proceedings of the fifth international AAAI conference on weblogs and social media*, 490-493, Menlo Park, California: The AAAI Press.
- Golder, S. A. and Macy, M. W. (2011) 'Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures', *Science* 333(6051), 1878-1881.
- González-Bailón, S., Borge-Holthoefer, J., Rivero, A. and Moreno, Y. (2011) 'The dynamics of protest recruitment through an online network', *Scientific Reports* 1, Article number 197, doi:10.1038/srep00197.
- González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J. and Moreno, Y. (2012) 'Assessing the bias in communication networks sampled from Twitter, [Online], Available: <http://ssrn.com/abstract=2185134>.
- Hoffmann, T. (2010) 'Netzgemeinde mobilisiert für Gauck', *taz.de*, [Online], Available: <http://www.taz.de/!53629/>.
- Jackson, N. and Lilleker, D. (2011) 'Microblogging, constituency service and impression management: UK MPs and the use of Twitter', *The Journal of Legislative Studies*, 17/1, 86-105.

- Jakat, L. (2010) “Astroturfing”—Geheimkampf um Botschaften im Netz’, *sueddeutsche.de*, [Online], Available:
<http://www.sueddeutsche.de/politik/streit-um-stuttgart-astroturfing-geheimkampf-um-botschaften-im-netz-1.1008550>.
- Jungherr, A. (2009) *The DigiActive guide to Twitter for activism*, [Online], Available: <http://andreasjungherr.net/wp-content/uploads/2011/10/Jungherr-2009-Digiactive-Guide-to-Twitter-for-Activism.pdf>.
- Jungherr, A. (2012) ‘The German federal election of 2009: The challenge of participatory cultures in political campaigns’, *Transformative Works and Fan Activism* 10, [Online], Available:
<http://journal.transformativeworks.org/index.php/twc/article/view/310/288>.
- Jungherr, A. and Jürgens, P. (2013) Forecasting the pulse: How deviations from regular patterns in online data can identify offline phenomena’, *Internet Research* 23(5): 589-607.
- Jungherr, A., Jürgens, P. and Schoen, H. (2012) ‘Why the Pirate Party won the German election of 2009 or the trouble with predictions: A response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. “Predicting elections with Twitter: what 140 characters reveal about political sentiment”’, *Social Science Computer Review* 30(2), 229-234.
- Jürgens, P. (2010) *Stell' Dir vor du twitterst und keiner hört zu. Themen und Öffentlichkeit auf Twitter*, Unpublished master's thesis at the Institut für Publizistik, Universität Mainz, Germany.
- Jürgens, P. and Jungherr, A. (2011) ‘Wahlkampf vom Sofa aus: Twitter im Bundestagswahlkampf 2009’, in Schweitzer, E. J. and Albrecht, S. (ed.) *Das Internet im Wahlkampf: Analysen zur Bundestagswahl 2009*, 201-225,

Wiesbaden: VS Verlag für Sozialwissenschaften.

Jürgens, P. and Jungherr, A. (forthcoming) ‘The use of Twitter during the 2009 German national Election’, *German Politics*.

Jürgens, P., Jungherr, A. and Schoen, H. (2011) ‘Small worlds with a difference: New gatekeepers and the filtering of political information on Twitter’, in *Proceedings of the ACM WebSci’11*, New York, NY, ACM. [Online], Available:

http://www.websci11.org/fileadmin/websci/Papers/147_paper.pdf.

Kleinberg, J. (2003) ‘Bursty and hierachical structure in streams’, *Data mining and knowledge discovery* 7(4), 373-397.

Kuhn, J. (2010) ‘Live aus der Baumkrone’, *sueddeutsche.de*, [Online], Available: <http://www.sueddeutsche.de/digital/stuttgart-protestvideos-im-netz-live-aus-der-baumkrone-1.1007082>.

Landmann, J. and Zuell, C. (2008) ‘Identifying events using computer-assisted text analysis’, *Social Science Computer Review* 26, 483-497.

Lang, K. and Lang, G. E. (1953) ‘The unique perspective of television and its effect: A pilot study’, *American Sociological Review* 18(1), 3–12.

Mader, F. (2010) ‘Schwabenstreich im Netz’, *sueddeutsche.de*, [Online], Available: <http://www.sueddeutsche.de/politik/protest-gegen-stuttgart-schwabenstreich-im-netz-1.998202>.

Maireder A. and Schwarzenegger, C. (2011) ‘A movement of connected individuals: Social media in the Austrian student protests 2009’, *Information, Communication & Society* 15(2), 171-195.

- Marwick, A. E. and boyd, d. (2011) 'I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience', *New Media & Society* 13(1), 114-133.
- Nikolov, S. (2012) *Trend or no trend: A novel nonparametric method for classifying time series (Master's thesis)*, Massachusetts Institute of Technology, Cambridge, MA.
- Petrovic, S., Osborne, M. and Lavrenko, V. (2010) 'Streaming first story detection with application to twitter', in *NAACL '10: Proceedings of the 11th Annual conference of the North American chapter of the association for computational linguistics*, 181-189, Stroudsburg, PA: ACL.
- Pfeiffer, T. (2010a) 'Stuttgart21 im Spiegel von Twitter', *web evangelisten*.
[Online], Available: <http://webevangelisten.de/stuttgart21-im-spiegel-von-twitter/>.
- Pfeiffer, T. (2010b) 'Stuttgart 21: Auf Facebook redet man nicht mit "den Anderen"', *web evangelisten*, [Online], Available:
<http://webevangelisten.de/stuttgart21-auf-facebook/>.
- Pickard, V.W. (2006) 'United yet autonomous: Indymedia and the struggle to sustain a radical democratic network', 28(3), 315-336.
- Reißmann, O. (2010) 'Riesenwut auf #S21-Polizeieinsatz', *Spiegel Online*,
[Online], Available:
<http://www.spiegel.de/netzwelt/web/0,1518,720701,00.html>.
- Sakaki, T., Okazaki, M. and Matsuo, Y. (2010) 'Earthquake shakes Twitter users: Real-time event detection by social sensors', in *Proceedings of the 19th international world wide web conference, WWW '10*, 851-860, New York, NY: ACM.

- Schimmelpfennig, M. (2010) 'Auf Twitter mehr bewegen: Ideen, den Widerstand zu stärken', *Copywriting*, [Online], Available: <http://copywriting.de/archives/1103>.
- Schrodt, P. A. (2010) *Automated production of high-volume, near-real-time political event data*, Paper presented at the Annual Meeting of the American Political Science Association, Washington, 2-5 September 2010.
- Segerberg, A. and Bennett, L. (2011) 'Social media and the organization of collective action: Using Twitter to explore the ecologies of two climate change protests', *The Communication Review* 14(3), 197-215.
- Shamma, D. A., Kennedy, L. and Churchill, E. F. (2009) 'Tweet the debates: Understanding community annotation of uncollected sources', in *Proceedings of the first SIGMM workshop on Social media, WSM '09*, 3-10, New York, NY: ACM.
- Shamma, D. A., Kennedy, L. and Churchill, E. F. (2011) 'Peaks and persistence: Modeling the shape of microblog conversations', in *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 355-358, New York, NY: ACM.
- Smith, A. (2011) *Twitter and social networking in the 2010 midterm elections*, Pew Internet & American Life Project, [Online], Available: <http://pewinternet.org/~media/Files/Reports/2011/PIP-Social-Media-and-2010-Election.pdf>.
- Smith, A., & Brenner, J. (2012) *Twitter Use 2012*, Pew Internet & American Life Project, [Online], Available: <http://pewinternet.org/Reports/2012/Twitter-Use-2012.aspx>.

- Stegers, F. (2010) 'The Revolution will be televised streamed via mobile', *onlinejournalismus.de*, [Online], Available:
<http://www.onlinejournalismus.de/2010/09/30/stuttgart-21-demo-the-revolution-will-be-televised-streamed-via-mobile/>.
- sueddeutsche.de (2010) 'Die ersten Bäume sind gefallen', [Online], Available:
<http://www.sueddeutsche.de/politik/protest-gegen-stuttgart-die-ersten-baeume-sind-gefallen-1.1006862>.
- Ternieden, H. (2010) 'Machtdemonstration gegen Mappus', *Spiegel Online*, [Online], Available:
<http://www.spiegel.de/politik/deutschland/0,1518,720840,00.html>.
- Vergeer, M., Hermans, L. and Sams, S. (2011) 'Is the voter only a tweet away? Micro-blogging during the 2009 European Parliament election campaign in the Netherlands', *First Monday*, 16/8, [Online], Available:
<http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3540/3026>.
- Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M., Schram, A. and Anderson, K. M. (2011) 'Natural language processing to the rescue? Extracting "situational awareness" tweets during mass emergency', in *Proceedings of the fifth international AAAI conference on weblogs and social media*, 386-392, Menlo Park, California: The AAAI Press.
- Weng, J. and Lee, B. (2011) 'Event detection in Twitter', in *Proceedings of the fifth international AAAI conference on weblogs and social media*, 401-408, Menlo Park, California: The AAAI Press.

White, J.S. and Klein, E. (2012) 'Coalmine: An experience in building a system for social media analytics', *Proc. SPIE 8408, Cyber Sensing 2012*, 84080A, doi:10.1117/12.918933.

Wienand, L. (2010) 'Mobiles Kamera-Einsatzkommando: Die "Volksreporter" von Stuttgart 21', *Rhein-Zeitung*, [Online], Available: http://www.rhein-zeitung.de/nachrichten/computerundmedia_artikel,-Mobiles-Kamera-Einsatzkommando-Die-Volksreporter-von-S21-_arid,151852.html.