3-6-2015

How Data Breaches Ruin Firm Reputation on Social Media! - Insights from a Sentiment-based Event Study

Griselda Sinanaj

Jan Muntermann

Timo Cziesla

How Data Breaches Ruin Firm Reputation on Social Media! – Insights from a Sentiment-based Event Study

Griselda Sinanaj, Jan Muntermann and Timo Cziesla

University of Göttingen, Göttingen, Germany
{griselda.sinanaj,muntermann,timo.cziesla}@wiwi.uni-goettingen.de

Abstract. Data breach events are heavily discussed in social media. Data breaches, which imply the loss of personal sensitive data, have negative consequences on the affected firms such as loss of market value, loss of customers and reputational damage. In the digital era, wherein ensuring information security is extremely demanding and the dissemination of information occurs at a very high speed, protecting corporate reputation has become a challenging task. While several studies have provided empirical evidence of the financial consequences of data breaches, little attention has been dedicated to the link between data breaches and reputational risk. To address this research gap, we have measured the reputational effect of data breaches based on social media content by applying a novel approach, the sentiment-based event study. The empirical results provide strong evidence that data breach events deteriorate the reputation of companies involved in such incidents.

Keywords: Data Breaches, Social Media, Reputational Risk, Sentiment Analysis, Sentiment-based Event Study

1 Introduction

In August 2014, the largest U.S bank J.P Morgan announced being victim of a data breach incident involving the theft of 76 million households data, including names, addresses, phone numbers and e-mail addresses, and 7 million data of small businesses [14]. Data breach incidents, which have become a common phenomenon businesses have to cope with, generate on average a unitary cost of $145 to the affected companies [28] and additional significant loss of market value to publicly listed firms [35].

Data breaches are negative events and as such, are also heavily discussed in social media. With this regard, survey studies reveal that data breach incidents have a negative influence on corporate reputation, due to the presence of social media venues acting as amplifiers through the rapid dissemination of information [9]. While research has primarily focused and provided empirical evidence of the financial implications of security incidents [35], apart from survey studies emphasizing the risk of reputational harm of data breaches, little attention has been devoted to the empirical investigation of the link between data breaches and corporate reputation.

12th International Conference on Wirtschaftsinformatik,
March 4-6 2015, Osnabrück, Germany

Sinanaj, G.; Muntermann, J.; Cziesla, T. (2015): How Data Breaches Ruin Firm Reputation on Social Media! - Insights from a Sentiment-based Event Study, in: Thomas. O.; Teuteberg, F. (Hrsg.): Proceedings der 12. Internationalen Tagung Wirtschaftsinformatik (WI 2015), Osnabrück, S. 902-916

To tackle the lack of research on this issue, we empirically investigate the reputation effect of data breach incidents by analyzing users' reaction on social media platforms. For this purpose, we measure the change in the social media sentiment before and after the disclosure of data breaches by applying a sentiment-based event study.

This study contributes primarily with new theoretical insights on the link between IT security incidents and corporate reputation and proposes a novel approach, which can be extended to other contexts. To the best of our knowledge, there have been no previous attempts of adapting the classical event study method to analyze event-related opinion formation utilizing unstructured data from social media.

The remainder of the paper is structured as follows. The following section focuses on the critical analysis of the current literature, establishes the research gaps and formulates the research questions. Further on, we proceed with the sample selection process and with a detailed description of the applied research methodology. We conclude with the interpretation of results and provide a summary of limitations and potential directions for future research.

2 Related Work

2.1 Economic Impact of Data Breaches

In spite of the substantial benefits to companies such as lower operating costs, increase of productivity and efficiency [25], technological advances have additionally increased the vulnerability of information systems. These systems are often target of skilled intruders who attack them and come into possession of large amounts of sensitive data [2], [34]. Data breaches belong to the broad category of information security incidents and imply the loss or theft of personal data records in electronic form, such as social security numbers, credit card numbers, user names and addresses [29]. The occurrence of such events can turn into large costs for the affected organizations. Tangible costs are immediately covered in the aftermath of the public dissemination of the incidents and include for instance, notification of customers through hotline customer support, forensic expertise [29], software and hardware costs, while intangible costs are not easily quantifiable and entail the loss of investor confidence, competitive advantage, trust and also reputational damage [35].

Several studies have attempted to quantify the intangible costs of information security breaches by measuring the financial impact of security breach announcements on listed companies [6 p. 68]. These studies adapt the event study method from the financial domain to measure the capital market reaction of the involved firms. This is the main aspect related to the economic impact of security breach events being actually addressed in the literature. There is yet no clear understanding of what are the dynamics of the other potential intangible costs, other than the loss of investor confidence and the financial impact. Hence, there are evident research gaps in the current literature that call for future contributions.

We claim that special attention should be especially devoted to the potential reputational losses originating from security breaches. Corporate reputation is commonly considered as one of the most valuable intangible assets that can help an organization

gain competitive advantage over its rivals. Although literature offers a wide range of definitions on corporate reputation [13], in this study reputation is defined as "the overall opinion about a firm by customers, investors, employees and the general public [7 p. 4].

2.2 Social Media and Corporate Reputation

With the advance of Web 2.0 technologies, social media has become an additional driver of reputation risk [3]. Content generated through communication in social media can become viral as it reaches and involves a large number of users worldwide [7]. While on the one hand, stakeholders such as consumers, investors and customers are free to post and exchange their personal thoughts and ideas on brands and products, on the other hand organizations do have little influence in terms of controlling or altering user generated content in social media [19]. As a consequence, companies do not have any more a full control on their reputation; in contrast, reputation in the social web environment is dictated primarily by the voice of users expressed in on line conversations [18].

Several cases show that social media can act as a reputational risk factor. For example, when an airline accidentally damaged a musician's guitar and refused to replace it, the musician published a music video about the incident on a social video platform. Millions of people watched the video and as the newspaper and television started to report on the story, the airline gave in, trying to prevent further reputational damage [3].

To measure corporate reputation on social media data, we make use of the sentiment analysis, whose widespread application in academia has coincided with the rise of the social media phenomenon and the consequent generation of large amounts of unstructured data. At the center of sentiment analysis or opinion mining is the study and analysis of humans' emotions, opinions and evaluations. Sentiment analysis has been applied in diverse research domains including finance and management [21], as well as for the measurement of corporate reputation based on social media content.

[31] used sentiment analysis to analyze the way social media content can be exploited for corporate reputation management. The study provides empirical evidence of the beneficial impact of social media on corporate reputation from a business agility perspective. [4] applied sentiment analysis to quantify reputation of ten large corporations based on historical social media data extracted from the microblogging platform Twitter. Sentiment analysis has been utilized to measure the daily sentiment values for each firm, while, in a second step, an estimating technique provides a linear representation of the sentiment values. Furthermore, the presented approach is validated by comparing the study results with the reputation ranking provided by the Reputation Institute, which is a survey-based and hence a classical measure of reputation. [7] developed instead an open source platform to measure online corporate reputation on the basis of real time Twitter stream data. Based on a predefined word lexicon, an algorithm generates a sentiment score for each incoming tweet based on the number of affective words. The open platform provides also a graphical representa-

tion of such sentiment values, which serve as a proxy for corporate reputation and offer hence the temporal evolution of the reputation values.

With respect to the measurement methods of corporate reputation, a very common approach used so far in academia relies on survey studies, e.g. the popular Fortune's survey of America's Most Admired Corporations. The participation on these surveys is usually reserved to a limited category of stakeholders such as the board of management and business analysts. Hence, such reputation measures do not encompass the evaluations and assessments of another category of important stakeholders, such as potential customers, consumers and employees [8]. Reputation measures anchored on social media data represent an alternative approach to survey-based measures which are obtained from the exploitation of unstructured data that differ substantially from the data at the basis of survey measures [4]. Corporate reputation measures based on social media content incorporate thus the opinions and assessments of a wider range of stakeholders.

In sum, prior research has already analyzed on the one side, the relationship between social media and corporate reputation and on the other side the financial consequences of information security incidents, while the interrelation of the triad data breach-social media-reputation has received little consideration. Hence, we aim at addressing this research gap by analyzing the reputational impact of data breach incidents on the basis of social media contents. In line with previous research, we apply the sentiment analysis method to analyze social media datasets in order to quantify corporate reputation [4], [7]. We separate ourselves from previous studies since we develop a new approach to investigate how corporate reputation is affected by critical discussions in social media following the public dissemination of data breaches. Based on this theoretical background we derive the following research questions that we aim to address in this study:

Research question 1 (RQ1): How to measure reputational effects of data breaches utilizing social media content?
Research question 2 (RQ2): Does social media promptly reflect newly available negative information on data breach incidents and how long does this effect persist?

3 Data Sources and Sample Selection

We exploited two distinct databases for the data collection process. The primary data set utilized for the sample selection comprises data breach incidents extracted from DataLossdb.org. This data source is an open-source relational database developed by the Open Security Foundation providing access to information on security vulnerabilities or data breaches. DataLossdb.org provides descriptive metadata about each single security incident recorded. Relevant to our study is the following information: incident ID, name of organization involved, date of occurrence and number of lost/stolen ID's [27]. Date of occurrence corresponds to the incident date as publicly (firstly) reported on primary news media sources, hereafter denoted as event date $[t_0]$.

The selection criteria applied to identify the final data sample are ranked as follows: a) Data breach incidents occurred worldwide between 1st January 2010 and 16th

November 2012. At this point, the dataset has been controlled for the presence of incident duplicates, which have been accordingly excluded. Multiple data breaches associated to the same firm were treated as separate events [1]. b) Since we seek to measure the reputational impact of data braches on global firms, the sample has been restricted to publicly traded companies at the event date. Incidents that affected privately held companies, governmental organizations, hospitals and universities have been accordingly removed. c) Finally, to ensure a significantly broad social media coverage for each security incident, we selected only those incidents with more than 30,000 lost or stolen records. Setting up this restriction on the original dataset increases the likelihood of extracting a high number of postings related to the specific firm involved in the data breach event.

Next we collected social media data on the sample of breach incidents through the social media monitoring tool SDL SM2. Social media data has been crawled from the following social media platforms: "blogs", "microblogs" (e.g. twitter), "social networks" (e.g. facebook), "online message boards", "wikis", "video- and photo-sharing" and "classified/review sites" [30]. To ensure the extraction of all postings referred to the company affected by the respective data breach, we set up the search query based on the company name e.g. "Citibank" and additionally refined the search by setting two parameters: English language and date range $[t_{-45}; t_5]$. Social media data has been collected for a total of 51 days, starting 45 days before the event date and 5 days afterwards, including the event date $[t_0]$ and entails only postings in English language. We provide further clarification on the date range parameter in the following methodological section. Searching through social media outlets based on the same search query along the entire time interval $[t_{-45}; t_5]$, aims at assuring consistency in the empirical analysis and avoiding biased results. The final output resulting from the search query contains the following relevant fields: result ID; media type (e.g. blog); author name; content of posting and timestamp (date of publication).

Further on, we screened thoroughly the data sample for the presence of confounding events (e.g. earning announcements or important managerial decisions) [23], whose disclosure overlaps with our predefined event window (-3;+5). This procedure although demanding, assures us that the effect on the reputation of the affected firms we are measuring is triggered from data breach announcements and not from such exogenous factors. Therefore, from the sample of breach incidents, two instances of data breaches have been discarded, leading to 40 data breach events.

Finally, we included only data breach events with postings on each single day within the time interval$[t_{-45}; t_5]$. This criteria (i.e. daily data) it is a necessary requisite to apply the sentiment-based event study approach. If the estimation or the event window contains any missing values, it would be necessary to apply interpolation techniques, which, in consequence, would counteract our objective to appropriately measure the changes in the sentiment values.

After applying the selection criteria we obtain a final sample of 30 data breach incidents. Table 1 provides a summary of our sample selection process.

Table 1. Sample selection criteria

Selection criteria	Number of observations left
Step 1: Data breach sample identification based on Datalossdb.org	
Time span: 1.1.2010 to 16.11.2012	1736
Public firms	282
Loss size: No. of ID's lost/stolen greater than 30,000	42
Confounding events during the event window	40
Step 2: Social media data collection based on SDL SM2	
Postings on each day within $[t_{-45}; t_5]$	30

4 Research Design

This section addresses our first stated research question RQ1 and contains a detailed explanation of the methodological approach applied in order to measure reputational consequences of data breaches based on social media content.

4.1 Sentiment Analysis

Three different methodologies can be utilized to perform sentiment analysis: linguistic, machine learning, and dictionary-based [7]. To analyze the content of texts extracted from social media platforms we apply the dictionary-based approach, which relies on predefined word lexicons for the text classification [11]. Social media data have been processed with the General Inquirer (GI) content analysis software based on the Harvard IV-4 psychosocial dictionary and is characterized by pre-labeled word lists with a particular semantic orientation such as positive, negative, strong, happy, sad. GI counts the word occurrences for the respective word category and provides a final output with the text classification [32]. Relevant to our context are the word categories labeled as "positive" and "negative". We use the sentiment polarity measure as a proxy for the overall users' opinion on social media [33 p. 1442]:

$$sentiment\ polarity = \frac{\#Words_{POS} - \#Words_{NEG}}{\#Words_{POS} + \#Words_{NEG}} \quad (1)$$

Sentiment polarity values are comprised in the interval [-1;+1], with the highest value of (+1) and the lowest value of (-1). In case the number of positive words equals the number of negative words, the text has a neutral sentiment and polarity is zero. The sentiment polarity variable will then be integrated into the sentiment-based event study approach in order to measure the reputational effects of data breaches.

4.2 Sentiment-based Event Study

To measure the impact of data breach announcements on corporate reputation, we combined the classic event study method [22] with the sentiment analysis. The theo-

retical fundament of the classical event study is the Efficient Market Hypothesis, which claims that any information newly available to the market will be instantly reflected by the asset prices [10]. Implementing an event study requires the specification of both an event window and an estimation window. While the estimation window is used for assessing price movements that can be expected when no significant event has happened, the event window covers an interval around the event date [22].

[26] have applied the event study method to analyze the effect of the Japanese banking crisis in non-financial companies considering a short estimation window of 40 days, between -60 and -20 days prior to the event under investigation. In line with previous research (e.g. [26]), we opted for a short estimation window comprising only 36 days starting at day -45 and ending at day -10. Communication through social media platforms generates large amounts of unstructured data whose extraction and processing for empirical research is time consuming [4]. Therefore, unlike many other classical event studies being based on long estimation windows (e.g. 250 trading days) we chose a relatively short one.

Next, we defined an event window of nine days covering the period from day -3 to day 5 including the event date t_0. The choice of this event window has two main purposes. First, the three days prior to the event date account for any leakage of information related to data breaches prior to its public disclosure through traditional media channels. In addition, it is common practice in classic event studies to consider event windows comprising several days following the event date, in order to observe the gradual recovery of stock prices after the information has been incorporated into the prices [16]. In doing so, we can observe at which point of time the effect of data breach disclosure will be entirely absorbed by the opinion formation observed in social media.

One disadvantage of a long event window, in particular if the analysis sample contains large corporations, is the high presence of other firm-related events or confounding events, whose effects blur the event study results [24]. If we opt for a longer window, it is very likely that we obtain a final sample of less than 30 data breaches, which in turn would reduce drastically the power of statistical tests [5]. In addition, previous studies using the classic event study method typically detect significant price effects only a few days prior and subsequent to the event dates [16].

In classical event studies, abnormal returns (AR_{it}) measure the deviation of the actual stock returns (R_{it}) from the ex-ante normal returns expected if the event did not occur [$E(R_{it}|X_{it})$] and are calculated as follows [22 p. 15]:

$$AR_{it} = R_{it} - E(R_{it}|X_{it}) \tag{2}$$

Cumulative abnormal returns (CAR_{it}) are obtained from the sum of AR_{it} for each day of the event window [22 p. 21]:

$$CAR_{it} = \sum_{t=1}^{N} AR_{it} \tag{3}$$

Following the above procedure, we define abnormal sentiment (AS_{it}) as the central variable of our sentiment-based event study in order to quantify the reputational effect of data breaches, formally specified as follows:

$$AS_{it} = S_{it} - E(S|X_{it}) \quad (4)$$

Since sentiment polarity S_{it} and $E(S|X_{it})$ assume values in the range [-1;1], AS_{it} values will be comprised in the interval [-2;2]. Similarly to the classical event study, cumulated abnormal sentiment (CAS_{it}) on each day of the event window is calculated with the expression:

$$CAS_{it} = \sum_{t=1}^{N} AS_{it} \quad (5)$$

There are basically three main statistical models used to evaluate the normal performance $E(R_{it}|X_{it})$ of stock prices in the context of event studies: (1) constant-mean return model which generates mean-adjusted returns; (2) market-adjusted return model and (3) the market model. We adopt the constant-mean return model based on the assumption that the ex-ante expected return $E(R|X_{it})$ of each security i is constant during the estimation window, i.e. $E(R|X_{it})= \mu$ [22]:

$$AR_{it} = R_{it} - \mu \quad (6)$$

The sentiment-based event study builds upon social media sentiment values and not on stock price returns unlike the classical event study. Adapting the market-adjusted return model would require the estimation of an overall market sentiment, which does not appear feasible. Furthermore, the market model is also not appropriate to our approach, as it requires the estimation of the market sentiment and of the model parameters. Hence, we measure abnormal sentiment AS_{it} values as follows:

$$AS_{it} = S_{it} - \mu \quad (7)$$

The constant-mean return model, although not being the most popular approach applied in practice, yields similar results as the other two models and does not influence the quality and the reliability of our empirical results. This is due to the low sensitiveness of the variance of abnormal returns against the normal returns model [5].

5 Empirical Analysis

5.1 Descriptive Statistics

Our final dataset comprises a sample of 30 data breach incidents and a total number of 388,635 postings obtained from social media platforms through the business intelligence software SDL SM2. The average number of records compromised by the breach incidents equals 4,350,237. The number of data records as well as the large number of postings illustrate respectively the relevance of our data breach sample and the richness of our social media dataset. With respect to breach source, for 70 % of the incidents the source of the breach is outside of the involved firms, for 23 % of the breaches inside of the firms and for the remaining 7% such information is not available. Considering the type of breach, 57% of the incidents were caused by hackers, 7% by fraud, followed by other types of breaches such as lost tapes, stolen drives and snail mails.

In terms of the distribution of data breaches over the time, recent years were characterized by a higher number of reported data breaches. The highest number of incidents were observed in 2012 (40%) and 2011 (40%), compared to 20% in 2010.

5.2 Results

To address our second research question RQ2, we computed over a nine-day long event window [-3;+5] the metrics average abnormal sentiment (AAS) and its cumulated effect over the event window, cumulated average abnormal sentiment (CAAS). To test the statistical significance of AAS and CAAS, we adopted and carried out the classic parametric *t*-test on the full sample. Formally, the validity of the null hypothesis $H_0 : \mu_{AS}(\mu_{CAS}) \geq 0$ has been tested against the alternative hypothesis $H_1 : \mu_{AS}(\mu_{CAS}) < 0$. Since the sample size equals 30 observations, the parametric approach is in this case applicable since the sampling distribution tends to be normally distributed [12 p. 134]. To test the robustness and the validity of results, we additionally report nonparametric test statistics based on the Wilcoxon signed-rank test. Mean and median values of AS and CAS are displayed for the event window along with the parametric- and nonparametric test results in table 2 and table 3 respectively. Mean AS (mean CAS) is equivalent to AAS (CAAS) respectively.

Table 2. Statistical test results on AAS between day (-3) and day (+5)

	Parametric test		**Nonparametric test**	
Day	Mean AS (% neg. AS)	*t*-statistic (*p*-value)	Median AS	*W*-statistic (*p*-value)
-3	0.029 (43)	1.184 (0.877)	0.019	278 (0.825)
-2	0.018 (43)	0.696 (0.754)	0.019	263 (0.735)
-1	-0.004 (53)	-0.163 (0.436)	-0.004	223 (0.428)
0	-0.178 (67)	-3.383 (0.001***)	-0.108	98 (0.002***)
+1	-0.152 (67)	-3.503 (0.001***)	-0.109	91 (0.001***)
+2	-0.095 (60)	-2.411 (0.011**)	-0.050	149 (0.044**)
+3	-0.123 (70)	-3.373 (0.001***)	-0.088	95 (0.002***)
+4	-0.083 (67)	-2.771 (0.005***)	-0.076	110 (0.005***)
+5	-0.067 (67)	-1.536 (0.068*)	-0.020	146 (0.038**)
*p<10%, **p<5%, ***p<1% (one-tailed test)				

Table 2 shows that both test procedures lead to similar results in terms of statistical significance, with exception of output values referred to day 5 (*p*-value=0.038 and *p*-value=0.068). Changes in sentiment polarity are detected from the day prior to the breach announcement (mean AS=-0.004, median=-0.004), supported also from the increase of 10% in the number of negative AS values (from day -2 to day -1). Substantial abnormal deviations of sentiment values from the normal performance are observed from the event day 0 until 5 days afterwards. The most negative value of AS is observed on the event day at which the incident became public. Additionally, from day -1 to day 0 we observe an increase of 14% of negative AS values, which provides

further evidence of the sensitiveness of social media users against data breach disclosures. *t*-test results on mean AS and Wilcoxon test results on median AS computed from day 0 to day 4, indicate strong statistical significance of the results at least at the 95% confidence level. On the last day of the event window, the results still remain significant despite the recovery of the downward trend of mean AS (median AS) values. In summary, we accept the validity of the alternative hypothesis H_1 and reject the null hypothesis H_0.

Table 3 summarizes CAAS values computed for each day of the event window. We observe that there is no negative sign of CAS in day -1, opposite to mean AS. The highest statistical significance at the 99% confidence level is achieved on days 1, 2, 3, 4, 5. Mean (median) CAS on the event date exhibit a lower significance of 90% (95%). These outcomes provide further evidence for the disapproval of the null hypothesis in favor of the alternative hypothesis. Hence, the abnormal sentiment is triggered by the security breach disclosure and is not simply attributable to pure chance.

Table 3. Statistical test results on CAAS between day (-3) and day (+5)

	Parametric test		**Nonparametric test**	
Day	Mean CAS (% neg. CAS)	*t*-statistic (*p*-value)	Median CAS	*W*-statistic (*p*-value)
-3	0.029 (43)	1.184 (0.877)	0.019	278 (0.825)
-2	0.046 (43)	1.151 (0.870)	0.008	268 (0.768)
-1	0.042 (50)	0.865 (0.803)	0.004	261 (0.722)
0	-0.136 (63)	-1.824 (0.039**)	-0.170	147 (0.040**)
+1	-0.288 (67)	-2.654 (0.006***)	-0.306	118 (0.009***)
+2	-0.384 (63)	-2.736 (0.005***)	-0.281	122 (0.011**)
+3	-0.507 (67)	-3.002 (0.003***)	-0.511	110 (0.005***)
+4	-0.590 (63)	-3.092 (0.002***)	-0.592	107 (0.004***)
+5	-0.656 (67)	-2.984 (0.003***)	-0.544	105 (0.004***)
*p<10%, **p<5%, ***p<1% (one-tailed test)				

Figure 1 depicts the behavior of the two variables, mean AS and mean CAS on each day of the event window. Data breaches reflect negatively on the reputational status of the involved firms, as evidenced by the sharp drop at the event date exhibited from both AAS (light-coloured line) and CAAS values (dark-coloured line).

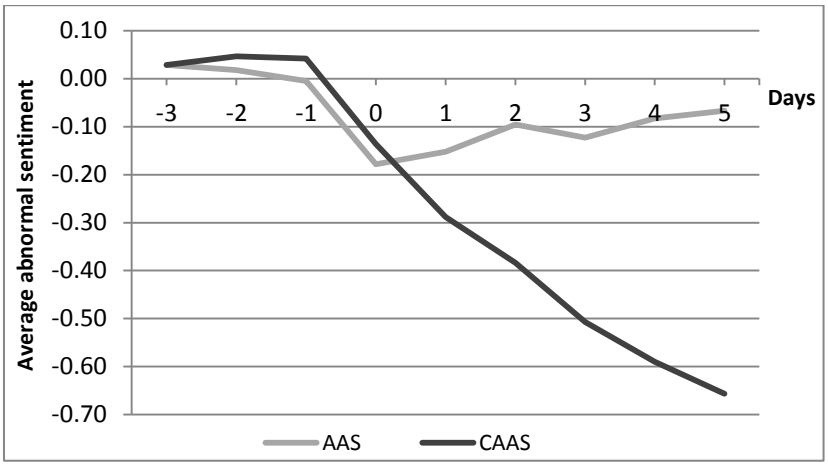

Fig. 1. AAS and CAAS values three days prior to the data breach disclosure and five days afterwards, including the event date [event window (-3;5)]

6 Discussion and Contributions

We observe statistically significant negative average abnormal sentiment (AAS) and cumulative abnormal sentiment (CAAS) following the public disclosure of data breaches until five days afterwards, which clearly show that such events effectively damage corporate reputation. The results do not reveal evidence of significant negative abnormal sentiment (cumulative abnormal sentiment) during the pre-event days, which indicates the absence of information leakage related to the breach incidents.

One major finding our study reveals is that the most negative values of AAS are observed immediately on the date of public disclosure of data breaches, meaning that the highest reputational losses occur when such events become of public domain. The figures evidence clearly how quickly the perception and the trust of individuals towards a company changes when their privacy sphere has been compromised. Such prompt reaction of the online community when data breaches are rendered public is however expectable due to the sensitive nature of the data type involved in data breaches. It is reasonable that people will lose trust and confidence on the involved firms as they blame them for such incidents and for not investing enough in information security [2]. On the one hand, the results implicate that privacy still remains a great concern for individuals, in spite of the frequent occurrence and coverage of data breaches in news media. On the other hand, organizations should adapt and apply appropriate security policies in order to prevent the future occurrence of such events.

Taking into account the processing of new information and the duration of the effect of data breaches in social media venues, the following observations can be made. With respect to users' behavior, similarly to the investors' reaction in the capital market scenario, stakeholders revise instantly the state of their beliefs, opinions, and evaluations based on the new flow of conveyed information, as evidenced by the abnormal change in the sentiment polarity. Hence, social media seem to process infor-

mation "efficiently", just as financial markets do under the condition of the efficient market hypothesis (EMH). In addition, reputational damage immediately observable from the public disclosure of data breaches indicates the vulnerability of this intangible asset under the influencing power of social media in the presence of negative events.

With focus on the duration of the data breach impact, the figures of AAS demonstrate significance from day 0 to day 5, implying that reputational damage caused by data breaches persists longer and is hence not entirely captured from the selected event window. This is a surprising finding since classic event studies have shown that asset prices typically reflect new information related to data breach announcements within a few days around the event date [1], [15]. Therefore, the results signalize that differences exist between investors' and social media users' behavior when considering the persistence of such reactions in time. Capital market reaction due to data breach announcements is observable in the short term within few days after the incident. Reputational harm is immediately observable in the aftermath of the event but unlike financial market reaction, lasts for a longer period of time. Hence, it would be interesting to observe and analyze the trend of reputational effects for a longer event window in order to obtain deeper insights on reputational effects in the long-term. This is of critical relevance to businesses, which have to respond to reputational damage with appropriate strategies in the long run.

With the findings of this study we contribute to the existent literature on the impact of information security incidents and raise new theoretical issues not tackled from previous research. Over the last decade, a body of IS literature has supplied empirical evidence of the financial implications of data breach incidents. Financial losses cover though one single aspect of the repercussions of such occurrences, which further encompass loss of trust and reputational tarnish. We contribute to this lack of knowledge with empirical results related to the intangible effects of data breaches and demonstrate thereby how data breach announcements damage firm reputation based on social media content and provide additional insights on the economic aspects of such incidents.

From a methodological perspective we provide a new approach stemming from the integration of sentiment analysis in the classical event study methodology. With the classical event study approach it is possible to quantify the change in the market value of firms involved in data breach incidents and measure therefore the financial impact triggered by such events on listed companies. With the sentiment-based event study, we provide instead a robust approach to measure the intangible effect of data breaches, such as reputational damage, and consequently contribute to the literature stream of corporate reputation. In addition, measurement approaches for the construct of corporate reputation proposed in the literature are mainly of qualitative nature and build upon large-scale survey studies, which are costly and require a long preparation time [4]. A clear advantage of our approach derives from the adoption of social media data to quantify corporate reputation, which entails valuable information that differs substantially from survey-based reputation measures.

The results emerged from this study have practical relevance for practitioners and businesses as well. As corporate reputation in the social web era is a primary concern

for every operating business, we provide evidence of the devastating effect of social media on corporate reputation when data breach incidents become known to the public. Hence, the insights of our study can help businesses to increase awareness on the risks social media can pose to corporate reputation in case of a negative scenario, and also, to promptly respond with the necessary strategic measures in order to protect their intangible assets. More generally, measuring social media sentiment on a continuous basis is a helpful instrument for businesses to monitor fluctuations of corporate reputation on real time. Hence, firms can exploit the variety of social media outlets to gain competitive advantage and online visibility in order to positively influence the overall online users' opinion and corporate reputation, too.

7 Limitations and Future Research

In spite of the theoretical and practical contributions, this study it is not exonerated from shortcomings and limitations. One limitation is the relative small sample size, mainly due to the fact that we are analyzing massive amounts of social media data, whose collection and preparation requires considerable efforts and a long processing time. Nevertheless, several studies on information security breaches are based on small samples too (e.g. [17] used 23 events while [20] used in their study 19 events).

Furthermore, the choice of a short event window comprising three days before the event date and only five days afterwards, hinders the observation and interpretation of reputational effects in the long term. A longer event window would contain a higher number of confounding events, whose esclusion would lead to a sample size of less than thirty data breaches, which would strongly influence the reliability of the statistical test results [5]. We aim though to address this limitation in our future research.

Collecting social media data based on a keyword search leads to the entity recognition problem, in the sense that we cannot control if the posting content is effectively directed to the specific firm or is has been simply mentioned by users in relation to another context. In addition, with regard to sentiment analysis the sentiment of words or sentences could be erronously classified as positive or negative, although the overall tone might be sarcastic, ironic or even without sentiment. These are some of the general problems when dealing with sentiment analysis techniques [21], which represent also a limitation of this study.

The findings of this study, although encouraging, constitute only the first step towards the investigation of the reputational impact and in general of the intangible effects of information security incidents, where further research is needed. Our goal is to extend this study by conducting a joint analysis of the classic- and sentiment event study with a larger data set and a longer event window. The contemporaneous analysis of the stock market, as well as of the social media reaction of data breaches, will provide a deeper understanding of the dynamics and the rationale behind opinion formation and investor behavior from a theoretical standpoint.

References

1. Acquisti, A., Friedman, A., Telang, R.: Is There a Cost to Privacy Breaches? An Event Study. In: Proceedings of the 21st International Conference on Information Systems (ICIS), pp. 1563–1580. Milwaukee, Wisconsin (2006)
2. Andoh-Baidoo, F.K., Amoako-Gyampah, K., Osei-Bryson, K.-M.: How Internet Security Breaches Harm Market Value. IEEE Security and Privacy. 8, 36–42 (2010)
3. Aula, P.: Social Media, Reputation Risk and Ambient Publicity Management. Strategy & Leadership. 38, 43–49 (2010)
4. Benthaus, J., Pahlke, I., Beck, R., Seebach, C.: Improving Sensing and Seizing Capabilities of a Firm by Measuring Corporate Reputation based on Social Media. In: Proceedings of the 21st European Conference on Information Systems (ECIS), pp. 1–12. Utrecht, Holland, (2013)
5. Brown, S.J., Warner, J.B.: Measuring Security Price Performance. Journal of Financial Economics. 8, 205–258 (1980)
6. Cavusoglu, H., Cavusoglu, H., Raghunathan, S.: Economics of IT Security Management: Four Improvements to current Security Practices. Communications of the Association for Information Systems. 14, 65–75 (2004)
7. Colleoni, E., Arvidsson, A., Hansen, L., Marchesini, A.: Measuring Corporate Reputation using Sentiment Analysis. In: Proceedings of the 15th International Conference on Corporate Reputation: Navigating the Reputation Economy, New Orleans, USA (2011)
8. Deephouse, D.L.: Media Reputation as a Strategic Resource: An Integration of Mass Communication and Resource-Based Theories. Journal of Management. 26, 1091–1112 (2000)
9. Economic Intelligence Unit.: IBM Global Reputational Risk and IT Study: How Security and Business Continuity can shape the Reputation and Value of your Company. Somers, NY (2012)
10. Fama, E.F., Fisher, L., Jensen, M.C., Roll, R.: The Adjustment of Stock Prices to New Information. International Economic Review. 10, 1–21 (1969)
11. Feldman, R.: Techniques and Applications for Sentiment Analysis. Communications of the ACM. 56, 82–89 (2013)
12. Field, A.: Discovering Statistics Using SPSS. SAGE Publications Ltd (2009)
13. Fombrun, C.J.: Reputation. Realizing value from the corporate image. Harvard Business School Press, Boston, MA (1996)
14. Forbes, http://www.forbes.com/sites/maggiemcgrath/2014/10/02/jp-morgan-says-76-million-households-affected-by-data-breach/ (Accessed: 27.10.2014)
15. Goel, S., Shawky, H.A.: Estimating the Market Impact of Security Breach Announcements on Firm Values. Information & Management. 46, pp. 404–410 (2009)
16. Goldstein, J., Chernobai, A., Benaroch M.: Event Study Analysis of the Economic Impact of IT Operational Risk and its Subcategories. Journal of Association of Information Systems. 12, 606–631 (2011)
17. Hovav, A., Arcy, J.D.: The Impact of Denial-of-Service Attack Announcements on the Market Value of Firms. Risk Management and Insurance Review. 6, 97–121 (2003)
18. Jones, B., Temperley, J., Lima, A.: Corporate Reputation in the Era of Web 2.0: the Case of Primark. Journal of Marketing Management. 25, 927–939 (2009)
19. Kaplan, A.M., Haenlein, M.: Users of the World, unite! The Challenges and Opportunities of Social Media. Business Horizons. 53, 59–68 (2010)

20. Ko, M., Dorantes, C.: The Impact of Information Security Breaches on Financial Performance of the Breached Firms: an Empirical Investigation. Journal of Information Technology Management. 17, 13–22 (2006)
21. Liu, B.: Sentiment Analysis and Opinion Mining. In: Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers (2012)
22. MacKinlay, A.: Event Studies in Economics and Finance. Journal of Economic Literature. 35, 13–39 (1997)
23. McWilliams, A., Siegel, D.: Event Studies in Management Research: Theoretical and Empirical Issues. Academy of Management Journal. 40, 626–657 (1997)
24. McWilliams, A., Siegel, D., Teoh, S.H.: Issues in the Use of the Event Study Methodology: A Critical Analysis of Corporate Social Responsibility Studies. Organizational Research Methods. 2, 340–365 (1999)
25. Mithas, S., Tafti, A., Bardhan, I., Goh, J.M.: Information Technology and Firm Profitability: Mechanisms and Empirical Evidence. MIS Quarterly. 36, 205–224 (2012)
26. Miyajima, H., Yafeh, Y.: Japan's Banking Crisis: An Event-Study Perspective. Journal of Banking & Finance. 31, 2866–2885 (2007)
27. Open Security Foundation, http://datalossdb.org/ (Accessed: 20.06.2014)
28. Ponemon Institute LLC. Cost of Data Breach Study: Global Analysis Benchmark research sponsored by IBM, Independently conducted by Ponemon Institute LLC (2014)
29. Romanosky, S., Telang, R., Acquisti, A.: Do Data Breach Disclosure Laws Reduce Identity Theft? Journal of Policy Analysis and Management. 30, 256–286 (2011)
30. SDL SM2 Social Media Monitoring, http://www.sdl.com/products/SM2/ (Accessed: 20.07.2014)
31. Seebach, C., Beck, R., Denisova, O.: Analyzing Social Media for Corporate Reputation Management: How Firms can Improve Business Agility. International Journal of Business Intelligence Research (IJBIR). 4, 50–66. doi:10.4018/ijbir.2013070104 (2013)
32. Stone, P.J., Dunphy, D.C., Smith, M.S., Ogilive, D.M.: The General Inquirer. The M.I.T. Press, Massachusetts (1966)
33. Tetlock, P.C., Saar-tsechansky, M., Macskassy, S.: More Than Words: Quantifying Language to Measure Firms' Fundamentals. The Journal of Finance. 63, 1437–1467 (2008)
34. Whitman, M.E.: In Defense of the Realm: Understanding the Threats to Information Security. International Journal of Information Management. 24, 43–57 (2004)
35. Yayla, A. A., Hu, Q.: The Impact of Information Security Events on the Stock Value of Firms: the Effect of Contingency Factors. Journal of Information Technology. 26, 60–77 (2011)