SECTION 2. Management in firms and organizations

Markus Stiglbauer (Germany), Anna-Lena Kühn (Germany), Christian Häußinger (Germany)

Semantic social media analytics of CSR image: the benefit to know stakeholders' perspective

Abstract

The issue of corporate social responsibility (CSR) and the use social media have risen in importance over the last few years. Companies are not solely reduced on their economic performance but also have to succeed ecologically and socially. Social media puts pressure on companies to do so and integrates different stakeholders' expectations towards and perceptions of a company with regard to CSR. Thus, social media are a suitable source for companies to match their CSR self image with their CSR public image. Data mining techniques and related software helps companies to assess their CSR public image and to adjust their CSR when there is a mismatch between self image and public image. This initiates a continuous CSR (control) cycle which helps companies to forecast accuracy of their firm value and to gain long-term competitive advantage against companies which do not manage and control their CSR strategically.

Keywords: data mining, social media, corporate social responsibility, firm value, firm reputation, stakeholder management.
JEL Classification: C80, G32, M14, M15.

Introduction

The 'responsibility' of companies has become an important issue in recent years. Nowadays, stakeholders do not only expect companies to succeed economically but also ecologically and socially (CSR). The main issue is that companies (should) reinvest in the general conditions of their own success, e.g. in terms of sustainable use of valuable resources. Stakeholders are affected by companies' activities or insensitivity for environmental and social problems. That is why 'responsibility' will not be just a temporary 'hot topic' (Porter & Kramer, 2006). Stakeholders are even able to exert pressure on companies to presuppose CSR. Consequently, companies develop strategies how to deal with increased CSR expectations (Eldomiaty & Choi, 2006). The CSR RepTrak™ 100 study shows that stakeholders' activities, such as investing or buying products are driven 60% by their perceptions of the company and only 40% by their perception of the products (Reputation Institute, 2012). Thus it is crucial for companies to focus not only on their self-image but to equally be aware of their public image/reputation. The critical reflection on how companies are doing business and what companies stand for seems crucial for their economic success.

Social media plays a significant role by being a communication tool that connects business and society. Through obtaining direct and rapid feedback, companies can better understand the expectations of their target groups. However, social media also bear risks, as resentment and disaffection can spread rapidly throughout the world. Non-governmental organizations, for example, use social media, such as Facebook or Twitter to point out disputable business activities to a broad public. Companies, in turn, experience difficulties not only in disproving allegations and reacting after a storm of criticism but also in evaluating the resulting economic consequences. Our study attempts to develop an analytical framework, in order to examine the impact of the CSR public image on economic success. Therefore, we analyze social media communication using data mining techniques to extract valuable information on the public image. This approach allows to detect and to understand patterns in social media communication. The inherent algorithms can be trained to use available data to generate predictions about unseen data. The synthesis between data mining and statistical simulation methods thus enables a semantic analysis of a high amount of input data. In a further step, we research inferences between the public image and the firm value as a key indicator for economic success.

Our paper is structured as follows. Following the introduction, section 1 briefly establishes the theoretical framework of CSR and CSR communication and illustrates the direct and indirect impact of CSR engagement. Section 2 highlights the technical process how social media data can be analyzed to detect patterns on CSR public image. Section 3 illustrates sample indications how CSR public image and firm value are connected. We finish with managerial implications and a conclusion to stress the importance of a continuous alignment of the CSR self-image with the public image.

1. Theoretical perspectives of CSR and CSR communication

1.1. CSR and CSR communication. In recent years, the role of business in society has changed and companies have taken on philanthropic responsibilities in areas like health, education, environment, infrastructure, and community development that were formerly borne mainly by governments. However, the concept of how companies should practice their corporate responsibility still remains fuzzy with unclear boundaries and debatable legitimacy (Lantos, 2001). So far, no universal definition of the CSR concept exists (Dahlsrud, 2008). Central statements on CSR are its rootedness in corporate ethical values, its link between economic interests and the environmental and social context and its consideration of the needs/concerns of society. A cross-country study shows the high CSR expectation towards companies: “set higher ethical standards and help build better society” (45%) vs. “make profit, pay taxes, create jobs, and obey the law” (8%) (Environics Research Group, 1999). In line with the growing awareness of CSR, the CSR communication has equally become more important. In contrast to the mandatory reporting of financial statements, the disclosure of CSR engagement is still a voluntary approach (Global Reporting Initiative, 2011). Nevertheless, the majority of society wants companies to openly discuss the addressing of social issues (Cone Inc., 2004). CSR communication can thus serve as a means to gain societal legitimacy (Brønn & Brønn, 2003) and to establish a positive public image/reputation (Hamann & Acutt, 2003). The key challenge is the communication of the CSR engagement without being accused of window dressing: Society is skeptical if everything seems too good to be true (Illia et al., 2013). Already since the 1980s, numerous scholars have attempted to clarify how CSR should be effectively communicated (Grunig, 1979; Manheim & Pratt 1986; Bruning & Lendingham, 1999; Dawkins, 2004). However, Internet has permanently changed the landscape of communication. Therefore, further investigation on its potential for CSR communication still remains necessary (Stuart & Jones, 2004; Kent et al., 2003). Despite the great use of social media (Dawkins, 2004), no study has specifically considered the importance of social media for CSR communication and CSR public image so far (Servaes & Tamayo, 2013).

1.2. Direct and indirect impact of CSR and CSR communication. Companies invest a lot in CSR (Reputation Institute, 2012; Rodgers et al., 2013) with direct and indirect effects for themselves. A well-elaborated CSR strategy can directly influence the business system, for example through setting CSR-oriented corporate goals in the supply chain management, making appropriate adjustments regarding energy efficiency, sustainably using resources in the production process or fostering research and development to execute eco-friendly operations (Sharma & Ruud, 2003). These initiatives are finally aimed at cost saving and risk reduction. This leads to a win-win situation for the environment, society and the companies. Yet, there are also indirect impacts that companies cannot control. These indirect consequences of CSR can have strong impacts on customer relations, companies’ public image or the information flow on social media channels (Weber, 2008).

By stating “one cannot not communicate” (Watzlawick et al., 1967, p. 48) companies are encouraged to realize that they always communicate either intentionally or unintentionally by everything they do or do not, report or do not report. In this context, it is important for companies to become aware of the significant differences between controllable direct corporate communications that comprise the corporate self-image and the uncontrollable, indirectly-induced public image. When companies directly moderate their CSR communication, they can benefit from official CSR reports, online presences on company websites, or through the registration of fans/followers on Facebook and Twitter (Kaplan & Haenlein, 2010). Such official communication channels promote the corporate self-image. In contrast, the public image of CSR arises through social media and is therefore called uncontrollable. These uncontrollable resources of social media belong, among others, to non-governmental organizations, news agencies, private persons and associations. This shift towards social media can be explained by the fact that users do not only receive information but also transmit information autonomously (Philippe & Durand, 2011). In doing so, they are capable of influencing the public image of companies. Due to a growing interest of society in CSR and its capability of understanding CSR (promoted by a widespread knowledge of this subject within the media and among opinion leaders) they do not solely depend on official CSR communication to evaluate CSR. By analyzing comments in social media, companies can become aware of their CSR public image. Through the deliberate use of such media, companies are involving in the opinion formation process and are thereby able to foster a bidirectional, dialogue-based and mutually beneficial stakeholder communication. With the aim of shedding light on the stakeholders’ perspective, we analyze social media communication by using data mining techniques. Thereby we distinguish three different kinds of unofficial and independent Internet resources: Facebook, Twitter and Google Alerts.

2. Semantic social media analytics of companies’ CSR image

2.1. RapidMiner, Google Alerts, Yahoo! Pipes and Social Media. In our study, data mining and qualitative semantic analysis are conducted with the program RapidMiner which automatically categorizes contributions and comments on Facebook and Twitter into positive and negative characteristics. Words with highly emotional connotations and important tokens aimed to classify text of social media represent the discriminants of the data mining process. By teaching the algorithm in the software to filter text, unknown comments and articles can be rated into a positive or negative intention (Kosorus et al., 2011). As shown in Figure 1, the cloud service Yahoo! Pipes processes web contents and is used in this study to combine different feeds connected with Google Alerts to only one feed. Consequently, RapidMiner requires only one operator to import feeds, whereby the computer performance can be reduced to make alternative use of it for processing. Since Yahoo! Pipes contains filters to separate relevant duties by Google Alerts, it filters terms like "corporate social responsibility" or "corporate citizenship" to identify duties about the CSR of the selected companies.

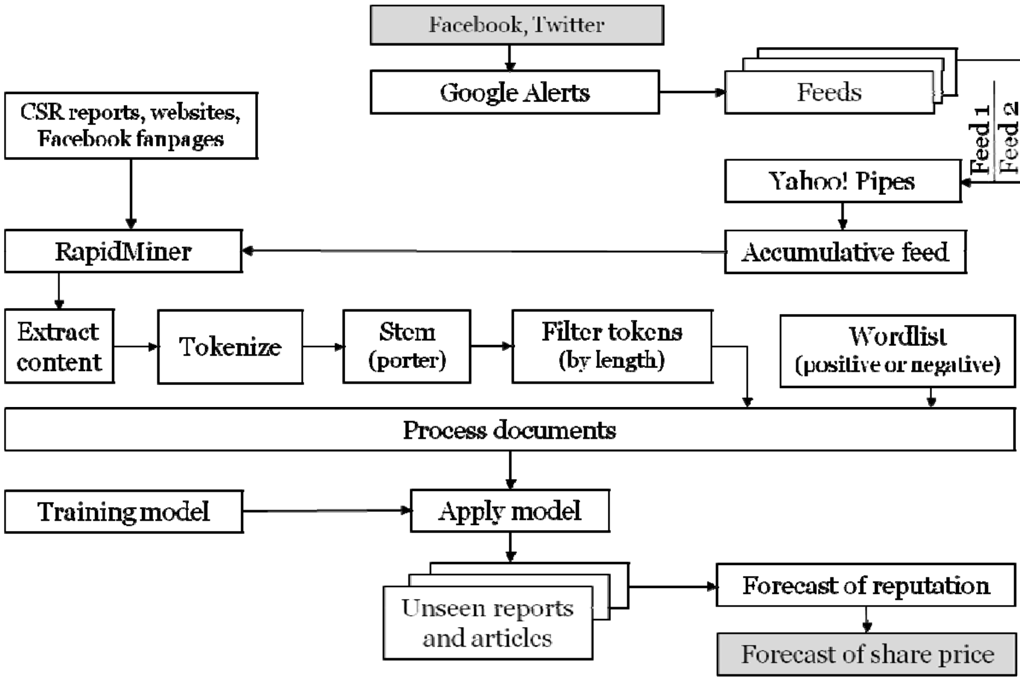

Fig. 1. Flowchart of the RapidMiner process

RapidMiner is programmed to check every half an hour after running the semantic web analysis, if an updated version of the feed is available and chooses one hundred random contributions of the combined feed. Due to different levels of corporate popularity, Google Alerts can find more Internet resources about certain companies. The more popular companies either belong to the business-to-consumer sector or are in the public eye due to public scandals, the financial crisis or preeminent marketing or CSR activities. As the combined feed weights articles about popular companies higher than other ones (Lin et al., 2006), the unofficial contributions are weighted by the average reported number of found Internet resources by Google Alerts to prevent distorted average weighting of the feed.

2.2. Applying the software. *2.2.1. Generating the word list.* Word lists are the basis of semantic analysis. They contain English words/phrases, which are generated in two different ways. First, a word list can be created by using a thesaurus. Second, the CSR reports within annual reports/on company websites can be used to generate them. The thesaurus produces words with positive and negative connotations, which are manually listed in a column in Microsoft Excel. Excel has a thesaurus function, but this function is normally applied only to a single word. Hence, we coded a macro program in Visual Basic for Applications, which applies the thesaurus function to an entire column of words as a pre-processing step (Nora et al., 2010). The macro program generates word synonyms for each cell. These synonyms may have positive or negative connotations compared to the original word. Thus, synonyms are only accepted in the word list if they are present in a list of synonyms with at least two primary labeled words. The original words are selected and labeled manually. If at least two of these words are the basis of two identical synonyms, the probability of finding synonyms with the same positive/negative label is similar to that of the original words. The newly generated synonyms are then used as the basis for the second round to identify further synonyms. The frequency of all words is counted after the second round. The output of the second round of synonym generation is usually identical to the original labeled words and the first round of synonym generation. If the same synonyms are generated repeatedly, their frequency is an indication of an accurate synonym label.

2.2.2. Processing the word list. The program has two components. The function of the first component is to initialize the algorithm before it classifies the text. The imported word list contains English words and phrases, which are classified based on their positive and negative meaning. The program operator transforms the word list into a collection of documents by outputting a document for each word or phrase in the word list. The newly created text objects serve as inputs for the next operator, which generates a term vector. The new dataset includes only a single data file. Therefore, the resulting data are specified by using a single text file. The tokens in the text are used to generate a word vector with the TF-IDF schema. Other sub-operators located within the word vector operators are used to process the word list.

2.2.3. Attribute vectors. The goal of this semantic analysis is to categorize all of the unofficial comments found in social media. Since the algorithm cannot understand connected text, we create word vectors as an interim step which include characteristic features like a string attribute. These features may include parts of the entire text with nominal attributes which form the basis of models that facilitate the comparison of pieces of text, while the values of string vectors can be equal or unequal. However, the string attributes contain no information about their textual relationships. The word vector contains the word frequency as a set of additional attributes, which is used to evaluate the meaning of a single word relative to the entire text (Suzuki, 2003). Initially, all of the characters in the word list are transformed to the lower or upper case. Selected regular expressions in tokens are then removed based on specified replacements. Next, an interior operator splits the word list into an array of tokens. As the splitting points are specified by the tokenization mode "non-letter character", a token consists only of a single word. Short English words that generally do not affect the text analysis, such as "is" or "a", are removed by the in-built stop word list operator. The next inner operator filters tokens that have a length of less than two characters.

All of the last five inner operators are part of the word vector operator, and they process the word list before creating the word vector. According to the word list, these operators are less important when processing social media content, but they are very helpful for analyzing phrases in the word list. The processed word list is used as an input for a similar operator in the second component of the program. The processed wordlist is employed as a guideline for classifying unknown text in social media content. The second output is a dataset that selects the attributes to be removed or those that should be part of the results. In this program, all attributes that do not constitute a missing value in any dataset are selected. The next operator utilizes the dataset as an input and changes the role of an attribute. The initial attribute is located in the word list column with the labels "positive" and "negative," where English words and phrases are classified (using a naive Bayesian classifier algorithm). This role is changed into the special attribute "label", which is used by the learning operator.

2.2.4. Training the algorithm. The learning operator includes a cross-validation process that estimates the performance of the algorithm. This cross-validation operator entails a testing and training sub-process. The inner testing operator splits the input dataset into a number of validation subsets. The training sub-process returns a model based on the input dataset. The testing sub-process generates a performance vector using the returned model while also quantifying its performance (Modha & Masry, 1998).

The performance measurement is an appraisal which is based on the final word list. Therefore, the performance may be lower when analyzing social media. The cross-validation operator has two outputs. It logs the performance and returns the average performance vector. Three other sub-operators function within the cross-validation operator. On the training side, a naive base learner outputs the classification algorithm based on the estimated normal distribution of the training dataset, using Laplacian correction to reduce the effects of zero probabilities (Zhang et al., 2006). On the testing side, an operator applies the model using the trained data, which is based on the imported word list. The trained data are used to classify the future contributions of social media. The other operator employed on the testing side is a performance operator, which considers the weight of the input data to calculate the performance of the applied model. It automatically detects the labeled input dataset, which contains two attributes: one with a role label and another with a role prediction. The second output of the cross-validation operator is the trained algorithm, which is used by the second function of the program for classifying text (Modha & Masry, 1998). The performance of the cross-validation learning operator can be improved by increasing the number of labeled words and phrases in the word list.

2.2.5. Cross-validation process. The cross-validation model evaluates the predictions of machine learning models to verify the categorizing attributed to the unseen data. Techniques, such as cross-validation, can also be used to compare the predictive performance of different machine learning methods. The requirements of this process are that all other factors having a possible influence remain unchanged. The cross-validation method divides the data into training and testing data. Typically, two-thirds are employed for training the learning algorithm, whereas one-third is used for testing the predictions of the known data (Modha & Masry, 1998). The presence of inappropriate proportions in the classes can lead to a one-sided learning bias, if many or all of the items in a specific class are not present in the two-thirds used as training data. To improve the quality of predictions, we need to select data mining methods with the lowest estimated error levels based on the results of cross-validation estimations.

2.2.6. Extracting relevant contributions from social media content. The second component of the program performs actual data mining of social media content. Initially, an Excel list that contains links to Facebook fanpages, Twitter feeds and forums about the stock market is imported. Each row in this input dataset contains a URL to an RSS feed. The next operator sends a GET request to each of the URLs and saves the pages temporarily using a page-specific attribute. The next two steps are similar to the first part of the program. One of the operators generates a collection of documents for each URL. The next operator uses the page tokens from the URL to produce a word vector. This operator has two inputs. The first is the list of labeled English words and phrases produced by the first component of the program. This word list is combined with the pages of the URL to generate a model based on a term vector in the following steps (Nora et al., 2010). The same sub-operators within this operator have to specify lower or upper case characters in a document to split documents into sequences of tokens and to filter the tokenized text based on English stop words and the length of words with less than two characters. There is only one other sub-operator that extracts the textual substance of the social media content from the HTML coding language (Hippner & Rentzmann, 2006).

2.2.7. Applying the algorithm. The final operator uses the trained algorithm to determine the “positive”/ “negative” intention of social media content. To achieve this, the operator uses the model and the labeled wordlist produced by the first component of the program. The unlabeled parts of the social media content are used as the second input dataset, which needs to be categorized. This operator joins both parts of the program, i.e., the labeled word list and the unlabeled contributions. The word list is one of the most important factors (Nora et al., 2010). The production of a larger number of labeled English words and phrases improves the ability of the learning operator when predicting unknown contributions in social media content. Both the word list and the algorithm are employed to analyze social media.

3. Sample indications and findings

We now demonstrate indications and findings when applying the above described semantic social media analytics of corporate CSR images. Based on a sample of 26 companies from the S&P 500 index, representing a cross section of the U.S. economy, the semantic social media analytics generated values for the CSR public image (0% worst; 100% best) and connected these to the firm value (share price multiplied by number of outstanding shares; a key indicator for economic success) for the 20 trading days of June 2012. To establish time series, RapidMiner requires an algorithm calculating the impact of the CSR public image on the share price development. Assigning explicit ID-numbers to the companies enables RapidMiner to connect share price data with the calculated image values (Diewald et al., 2008). Subsequently, the multivariate time-series, consisting of share prices and public image values, are analyzed by an operator within RapidMiner.

It transforms all data into single, daily accrued information. Furthermore, the share price is determined as a forecast variable by adjusting share price of the previous to the next trading day (Morik et al., 2010). The forecast share price is then compared to the opening price on the next trading day. When both values move into the same direction, RapidMiner calculates a forecast model based on the trend of the image value in connection with the share price. This algorithm improves its forecast accuracy the more daily share prices and image values are included. In the further process of the training model the estimate algorithm is validated.

This validation implies a comparison of the forecast share prices with the opening prices. The quality of the forecast estimator is then determined by consistent share price developments. After transforming the forecast prices into vectors by scaling them into values between -1 and +1 the single values are temporally ordered in the form of point estimates (Modha and Masry, 1998). Finally, the point estimates are converted into the algorithm to create a time series conducting Monte Carlo simulation with R (GNU S). In the ideal case, the forecast time series runs parallel to the time series of the opening prices. The forecast estimator is evaluated in an iterative process by checking each group of five forecast share prices on its forecast accuracy 20 times in a sequence.

Table 1 shows the output of these calculations for WellPoint. Table 2 summarizes the results for the full sample.

Table 1. Forecast value incl./excl. CSR public image for WellPoint (first/last trading day per week stated)

Date	CSR public image	Forecast *incl.* CSR public image	Forecast *excl.* CSR public image	Share price	Forecast development	Market development	Forecast accuracy	Expected gain/loss
06/04/2012	83%	66.85	66.89	65.45				
06/08/2012	83%	68.93	69.05	69.07	Down	Down	1	1.80
06/11/2012	55%	69.07	69.01	69.01	Up	Up	1	0.06
06/15/2012	83%	69.91	70.06	70.77	Down	Down	1	0.97
06/18/2012	83%	70.24	70.40	71.34	Down	Down	1	0.57
06/22/2012	80%	69.35	69.46	69.76	Down	Down	1	0.30
06/25/2012			68.98	68.95	Up	Up	1	0.81
06/25/2012	83%	68.86		68.95	Down	Up	0	-0.81
06/29/2012	83%	65.90	65.90	63.79	Up	Up	1	2.11
Forecast value incl. reputation value							73.68%	7.68
Forecast value excl. reputation value							78.95%	9.30

Table 2. Forecast value and expected gains/losses incl./excl. CSR public image

Company	Expected return *incl.* public image	Expected gain/loss	Expected return *excl.* public image	Expected gain/loss	Difference
Alcoa	73.68%	0.78	68.42%	0.60	+0.18
Altria Group	63.16%	2.62	63.16%	2.62	+/- 0
Bank of America	68.42%	1.84	73.68%	1.88	-0.04
Best Buy	84.21%	7.62	78.95%	4.96	+2.66
CarMax	68.42%	3.64	63.16%	2.48	+1.16
Chevron	68.42%	9.10	73.68%	11.82	-2.72
Clorox	63.16%	2.04	68.42%	2.52	-0.48
Dean Foods	57.89%	1.63	63.16%	1.65	-0.02
Disney	68.42%	3.67	63.16%	2.95	+0.72
ExxonMobil	68.42%	5.44	63.16%	4.18	+1.26
General Dynamics	68.42%	8.60	63.16%	4.16	+4.44
Goldman Sachs	78.95%	17.82	73.68%	13.44	+4.38
IBM	73.68%	26.66	73.68%	28.64	-1.98
Johnson & Johnson	73.68%	4.92	78.95%	6.02	-1.10
JPMorgan Chase	63.16%	4.73	63.16%	4.73	+/- 0
McDonald's	52.63%	1.71	57.89%	3.01	-1.30
Nike	73.68%	22.72	73.68%	22.72	+/- 0
Procter & Gamble	47.37%	2.23	57.89%	0.15	+2.08
Pfizer	73.68%	3.39	73.68%	3.39	+/- 0
Starbucks	47.37%	2.23	68.42%	8.46	-6.23
Coca-Cola	73.68%	7.67	73.68%	7.03	+0.64
Dow Chemical	68.42%	2.75	73.68%	3.81	-1.06
Walmart	66.67%	6.41	55.56%	-1.01	+7.42
WellPoint	73.68%	7.68	78.95%	9.30	1.62
Wells Fargo	73.68%	3.39	73.68%	3.39	+/- 0
Yum!	77.78%	6.56	66.67%	7.02	-0.46
Portfolio development		11.82%		10.86%	

We can conclude that for 81% of the companies, the forecast on the expected gain/loss including the CSR public image did show a divergence. Obviously the single differences seem to be small at first sight. Still, the performance of the portfolio including the CSR public image is about 1% higher than the time series portfolio. This indicates an inference between the CSR public image and the firm value. Hence, the monitoring of the public image enables companies to improve the forecast accuracy of their firm value.

Implications and conclusions

CSR is at the top of the agenda of companies, governments, supranational organizations and the global society (Moreno & Capriotti, 2009). The appropriate relationship between business and society has become the focus of the debate (Schwartz & Carroll, 2003). This debate goes beyond the real world and is spread out in the virtual world of social media (Brønn & Brønn, 2003). CSR engagement gives occasion to a continuous dialogue between companies and stakeholders: "A dialogic loop allows publics to query [companies] and, more importantly, it offers [companies] the opportunity to respond to questions, concerns and problems" (Kent & Taylor, 1998, p. 326). Based on our findings, we draw the conclusion that companies should work in close collaboration with their stakeholders – aiming to maximize the shared value created for stakeholders and society (European Commission, 2011). Since it is generally acknowledged that CSR can become a competitive advantage (Porter & Kramer, 2006), it seems indispensable for companies to get insight into the public perception of their social responsibility engagement (Figure 2).

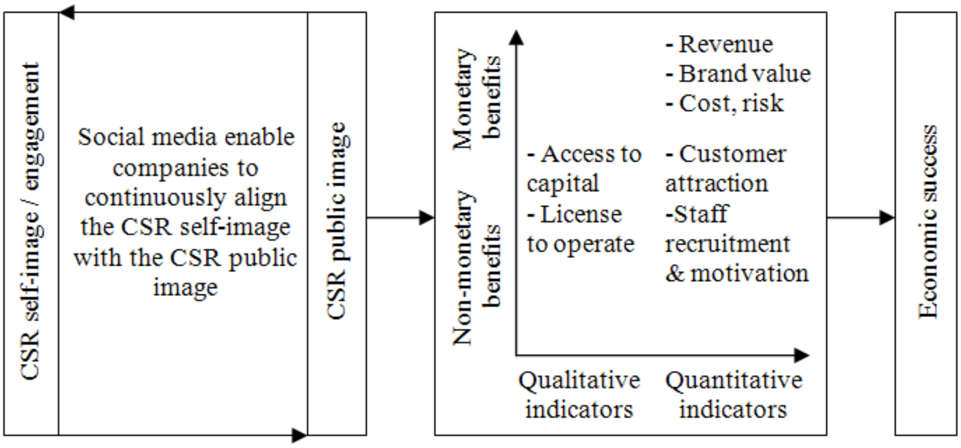

Fig. 2. Social media enable the alignment of the CSR self- and public image

Managers in charge of CSR issues need to consider the strategic impact (Podnar, 2008) and evaluate the success of their CSR projects (Kaplan & Norton, 2008), in order to make the right decisions on their future CSR engagement. We have demonstrated that extracting the CSR public image out from social media helps companies to predict their firm value. Nevertheless, we suggest that companies engaging in and reporting on their CSR engagement can only add value, if the CSR self-image/engagement and the public image are aligned. Hence, companies with a poor CSR public image are unlikely to reap any immediate benefits from engaging in CSR. In fact, such engagement may appear disingenuous and may well have the opposite effect. In the long run, the engagement in and dissemination of such engagement could create value, if companies change their CSR public image (Servaes & Tamayo, 2013).

Through the interaction with diverse stakeholders companies are integrated into an ongoing social participation process (Clark, 2000): Business in society is not just about what is going on in business, it is about what is going on in society, too. Through social media, companies gain a following of people who are interested in their CSR engagement and can keep tabs on stakeholder sentiment in any emerging issue (Mohin, 2012). This responsiveness (one foot inside and one foot outside the company-approach) of companies to society's agenda will be a contribution and a key determinant in shaping the companies, markets and society of the future (Fitzgerald & Cormack, 2006).

References

1. Brønn, P.S. and Brønn, C. (2003). A reflective stakeholder approach: Co-orientation as a basis for communication and learning, *Journal of Communication Management*, 7 (4), pp. 291-303.
2. Bruning, S.D. and Lendingham, A. (1999). Relationships between organizations and publics, development of a multi-dimensional organization-public relationship scale, *Public Relations Review*, 25 (2), pp. 157-170.
3. Carroll, J. (2005). *The Blacksburg Electronic Village: A study in community computing*. In: Van den Besselaar, P. and Kiozumi, S. (eds.), Digital Cities III: Information Technologies for Social Capital. Springer-Verlag, New York, pp. 43-65.
4. Clark. A. (2000). They're talking about you: Some thoughts about managing online commentary affecting corporate reputation, *Journal of Communication Management*, 5 (3), pp. 262-276.
5. Cone Inc (2004). *Cone corporate citizenship study*, Cone Inc, Boston, MA.
6. Dahlsrud, A. (2008). How corporate social responsibility is defined: An analysis of 37 definitions, Corporate Social Responsibility and Environmental Management, 15 (1), pp. 1-13.
7. Dawkins, J. (2004). Corporate responsibility: The communication challenge, *Journal of Communication Management*, 9 (2), pp. 108-119.

8. Diewald, N., Stührenberg, M., Garbar, A. and Goecke, D. (2008). Serengeti-Webbasierte Annotation semantischer Relationen, *Journal for Language Technology and Computational Linguistics*, 23 (2), pp. 74-90.
9. Eldomiaty, T.I. and Choi, C.J. (2006). Corporate governance and strategic transparency: East Asia in the international business systems, *Corporate Governance: International Journal of Business in Society*, 6 (3), pp. 281-296.
10. Environics Research Group (1999). *Millennium poll on corporate social responsibility*, Environics International Ltd., Toronto.
11. European Commission (2011). *More responsible businesses can foster more growth in Europe*. European Commission, Brussels.
12. Fitzgerald, N. and Cormack, M. (2006). *The Role of Business in Society*, CSRI Report No. 12.
13. Global Reporting Initiative (2011). *Sustainability reporting guidelines, Version 3.1*. The Global Reporting Initiative, Amsterdam.
14. Grunig, J.E. (1979). A new measure of public opinions on corporate social responsibility, *Academy of Management Journal*, 22 (4), pp. 738-764.
15. Hamann, R. and Acutt, N. (2003). How should civil society (and the government) respond to 'corporate social responsibility'? A critique of business motivations and the potential for partnerships, *Development Southern Africa*, 20 (2), pp. 255-270.
16. Hippner, H. and Rentzmann, R. (2006). Text mining, *Informatik-Spektrum*, 29 (4), pp. 287-290.
17. Illia, L., Zyglidopoulos, S.C., Romenti, S., Rodríguez-Cánovas, B. and González del Valle Brena, A. (2013). Communicating corporate social responsibility to a cynical public, *Sloan Management Review*, 54 (3), pp. 16-18.
18. Kaplan, A.M. and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media, *Business Horizons*, 53 (1), pp. 59-68.
19. Kaplan R.S. and Norton, D.P. (2008). Mastering the management system, *Harvard Business Review*, 86 (1), pp. 1-17.
20. Kent, M. and Taylor, M. (1998). Building dialogic relationships through the worldwide web, *Public Relations Review*, 24 (3), pp. 321-334.
21. Kent, M., Taylor, M. and White, W. (2003). The relationship between Web site design and organization responsiveness to stakeholders, *Public Relations Review*, 29 (1), pp. 63-77.
22. Kosorus, H., Hönigl, J. and Küng, J. (2011). *Using R, WEKA and RapidMiner in time series analysis of sensor data for structural health monitoring*. In: IEEE Computer Society (ed.) Proceedings of the 2011 22nd International Workshop on Database and Expert Systems Applications, IEEE Computer Society, Washington DC, pp. 306-310.
23. Lantos, G.P. (2001). The boundaries of strategic corporate social responsibility, *Journal of Consumer Marketing*, 18 (7), pp. 595-630.
24. Lin, Z., Li, D., Janamanchi, B. and Janamanchi, W. (2006). Reputation distribution and consumer-to-consumer online auction market structure: An exploratory study, *Decision Support Systems*, 41 (2), pp. 435-448.
25. Manheim, J.B. and Pratt, C.B. (1986). Communicating corporate social responsibility, *Public Relations Review*, 12 (2), pp. 9-18.
26. Modha, D.S. and Masry, E. (1998). Prequential and cross-validated regression estimation, *Machine Learning*, 33 (1), pp. 5-39.
27. Mohin, T.J. (2012). *Changing business from the inside out*. Berrett-Koehler Publishers, San Francisco, CA.
28. Moreno, A. and Capriotti, P. (2009). Communicating CSR, citizenship and sustainability on the web, *Journal of Communication Management*, 13 (2), pp. 157-175.
29. Morik, K., Deuse, J., Faber, V. and Bohnen, F. (2010). Data Mining in Sensordaten verketteter Prozesse – Ein Ansatz zur autonomen Automation in komplexen, verketteten Produktionsprozessen mit Hilfe maschineller Lernverfahren, *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 15 (1-2), pp. 106-110.
30. Nora, B., Lemnaru, C. and Potolea, R. (2010). *Semi-supervised learning with lexical knowledge for opinion mining*. In: IEEE Computer Society (ed.) Proceedings of the 2010 IEEE 6th International Conference on Intelligent Computer Communication and Processing, IEEE Computer Society, Washington DC, pp. 19-25.
31. Philippe, D. and Durand, R. (2011). The impact of norm-conforming behaviors on firm reputation, *Strategic Management Journal*, 32 (9), pp. 969-993.
32. Podnar, K. (2008). Communicating corporate social responsibility, *Journal of Marketing Communications*, 14 (2), pp. 75-81.
33. Porter, M.E. and Kramer, M. (2006). Strategy & society: The link between competitive advantage and corporate social responsibility, *Harvard Business Review*, 84 (12), pp. 76-92.
34. Reputation Institute (2012). *2012 CSR RepTrak™ 100*, Reputation Institute, New York, NY.
35. Rodgers, W., Choy, H.L. and Guiral, A. (2013). Do investors value a firm's commitment to social activities? *Journal of Business Ethics*, 114 (4), pp. 607-623.
36. Schwartz, M.S. and Carroll, A.B. (2003). Corporate social responsibility: A three-domain approach, *Business Ethics Quarterly*, 13 (4), pp. 503-530.
37. Servaes, H. and Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness, *Management Science*, 59 (5), pp. 1045-1061.
38. Sharma, S. and Ruud, A. (2003). On the path to sustainability: Integrating social dimensions into the research and practice of environmental management, *Business Strategy and the Environment*, 12 (4), pp. 205-214.
39. Stuart, H. and Jones, C. (2004). Corporate branding in marketspace, *Corporate Reputation Review*, 7 (1), pp. 84-98.

40. Suzuki, S. (2003). *Probabilistic word vector and similarity based on dictionaries*. In: Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Text Processing, Springer-Verlag, Berlin/Heidelberg, pp. 562-572.
41. Watzlawick, P., Beavin, J.H. and Jackson, D.D. (1967). *Pragmatics of Human Communication*, Norton, New York, NY.
42. Weber, M. (2008). The business-case for corporate social responsibility: A company-level measurement approach for CSR, *European Management Journal*, 26 (4), pp. 247-261.
43. Zhang, J., Kang, D.K., Silvescu, A. and Honavar, V. (2006). Learning accurate and concise naive Bayes classifiers from attribute value taxonomies and data, *Knowledge and Information Systems*, 9 (2), pp. 157-179.