Presented under the title:

**Measuring Corporate Reputation using Sentiment Analysis**

Elanor Colleoni  
Centre for CSR, Copenhagen Business School, Denmark, elc.ikl@cbs.dk

Adam Arvidsson  
Department of Political and Social Sciences, University of Milano, Italy,  
adam.arvidsson@unimi.it

Lars K. Hansen  
Cognitive Systems, Technical University of Denmark

Andrea Marchesini  
Senior Engineer at Mozilla Corporation, Copenhagen, Denmark
Monitoring Corporate Reputation in Social Media using Real-time Sentiment Analysis

Elanor Colleoni, Adam Arvidsson, Lars K. Hansen and Andrea Marchesini

Abstract

In recent years, new digital media have become important for social networking and content sharing. Due to their large diffusion, social media platforms have also both increased the strategic importance of managing corporate reputation and rendered this more difficult. Companies are increasingly apprehensive about information and opinions that can spread through online communities rapidly without any control. While social media platforms increase the power of stakeholders, they also represent a large-scale source of information about feelings, opinions and sentiments of people that allow us to measure and monitor reputation through the analysis of user generated content in real-time. In this paper, we show how social media content can be used to measure the online reputation of a company. Furthermore, we present an open platform that uses a sentiment analysis algorithm on twitter traffic to monitor the real time evolution of company reputation.

Introduction

Corporate reputation has become a strategic issue for management and companies (Fombrun, 2001). Companies have understood to capitalize on their reputation in valuing intangible assets and attracting financial capital, and knowledge workers depend on their personal brands or other reputational assets to set the market price of their skills and talents.
To a great extent this new interest in reputation has been triggered by the new dynamics of public communication that have followed from the diffusion of networked digital media, and particularly of social media like Twitter and Facebook that have created a fundamentally different situation for reputational management. Social media are characterized by “easy searching, open participation, a minimal publishing threshold, dialogue, community networking and the rapid and broad spread of information and other content via a wide range of feedback and linking systems” (Pekka, 2010: p.44). In the last few years, social media platforms have dramatically increased their capacity to capture online attention. According to recent statistics from Nielsen Company, sites like Facebook and Twitter now account for 22.7% of time spent on the web (Nielsen 2010). Apart from their great potential as new channels for commerce, such as viral marketing, the impact of these new technologies is often stated in terms of reputational risk in academic and practitioners articles (Gorry & Westbrook, 2009; Gaines-Ross, 2010; Pekka, 2010). For example, Gorry & Westbrook (2006) reported the case of the AOL, where a customer recorded a session with an arrogant and unresponsive customer service representative, posted it on YouTube and made AOL an Internet laughing stock. Pekka (2010) has described the reputation loss of a car dealership in Finland after the story of an insulted customer was spread using a chat room. However, there is a lack of empirical investigation in the new possibilities offered by the availability of online consumer opinions and sentiment for reputation management. Indeed, with the massive data produced in social media sites, it is possible to extract, monitor and even predict corporate reputation trends by aggregating subjective opinions using data mining techniques, such as opinion mining and sentiment analysis (Ogneva, 2010).
In this paper we propose a new measure of online corporate reputation based on the analysis of affective flows in social media using sentiment analysis. Furthermore, we present a tool for monitoring real-time corporate reputational trends on Twitter.

**Corporate Reputation in Online Social Media**

The development of an information economy, and in particular its more recent 'social economy' phase, has been the *pluralization of conceptions of value* (Stark, 2009). A number of business actors are discovering that satisfying such alternative orders of value can provide important business opportunities both in the short and in the long run. The rise of brands; the growing importance of reputation, both for companies and individuals, and the weight that perceptions of social responsibility and ethical conduct has on consumers, employees, investors and other stakeholders are all manifestations of this. Many scholars have set the strategic importance of these intangibles in creating market barriers and strengthening competitive advantages. In other words the ability to create, manage and exploit these intangibles, in the firms’ perspective allows them to drive markets rather than to be market-driven. Despite the fact that corporate reputation is become a central issue for management and strategic marketing studies, substantial difference exists in its definition (Fombrun, 1996; Fombrun & Van Riel, 1997; Wartick, 2002). Rindova et al. (2005) have reviewed over 60 articles from six different journals and have identified two different perspectives when studying reputation.

But what is reputation? From an economics perspective, reputation is characterized by the particular attributes of a firm and its past performances; whereas scholars from institutional theory tend to define it as a collective knowledge about a firm shared by stakeholders (Fombrun,
1996). The resulting definition of organizational reputation is comprised by two dimensions: stakeholders’ perceived quality and organization prominence in the minds of stakeholders (Rindova et al., 2005). Stakeholders’ evaluative perceptions can be described as the overall opinion about a firm by customers, investors, employees and the general public, which is expressed as an attitudinal construct, where attitude denotes subjective, emotional, and cognitive based mindsets (Hall, 1992). The organization prominence refers to the degree of large-scale collective attention and recognition (Einwell et al. 2010). Thus, corporate reputation can be seen as “the result of a public judgment that increases (or decreases) over the time and it is socially shared by different stakeholders” (Siano et al., 2010: p.6). There is ample documentation showing that social media and Web 2.0 have strengthened the role of company reputation and brand value in a wide variety of economic decision making processes: consumers consult social media based reputation in their choice of brands, talented employees are sensitive to social media based reputation in deciding to exit or remain with a company and investors increasingly use analysis of social media sentiment as part of their investment decisions. Digital media has reshaped the way that organizations gain both recognition and affective attachment from their public.

Large-scale Collective recognition

The growing importance of reputation management and corporate social responsibility is directly related to the diffusion of a media culture, and its penetration within a multitude of social relations, like those between buyers and sellers, or investors and companies. This has meant that the ability to construct a positive aura- a good reputation, a positive affective climate, an attractive identity a 'good feeling'- has become crucial to business success (Hunt, 2007).
Arguably, the diffusion of web 2.0 has both increased the strategic importance of managing corporate reputation and has rendered this more difficult. This latter is to a large extent due to the diffusion of digital technologies that have shaped the nature of organization communication and the consumer empowerment that this has entailed. Traditionally, public opinion was thought to be channeled by a number of important media institutions, such as newspapers, radio and television stations, and corporate communication was to a large extent conceived as the practice of using these institutions in order to convey a desired message about a company. This model implied a fair amount of freedom for corporate communication vis a vis and an audience that had little agency in defining the truthfulness of such communication. Today academics and professionals tend to suggest that public opinion has grown more independent in relation to media institutions and companies. Indeed, communication in social media tends to be mostly viral, participatory and bi-directional, and consequently opinions about companies are more difficult to control.

**Viral**

The concept of virality “refers to the potential of unstructured social relations like gossip, word of mouth, and lately online sociality to function as a medium of communication” (Hansen et al. 2011: p.2). The idea is that highly satisfied or unsatisfied customers are very likely to share their experiences within their social networks, like among friends or colleagues, and in so doing they influence the perception of potential customers about a company (Reichheld, 2003). During the last fifty years, sociologists and communication theorists have shown that people are more likely to trust and consequently to behave according to the general opinions that circulate in their social networks (Kats & Lazarsfeld, 1955). Before the diffusion of Internet, social networks were based
on face-to-face relationship, geographically located and obviously quite small. The only way to
spread a message to a large audience was by using mass-media systems, such as newspapers and
radio. This implied that the voice of the customers was overwhelmed by the mass-media voice.
Indeed, traditional broadcast media are based on a hierarchical one-to-many communication,
with a clear distinction between producer and consumer of information and an audience that does
not participate in the creation and selection of content. The advent of new digital social media
has completely changed the picture. Digital media are organized in the same way as offline
social networks where non-hierarchical communication flows in decentralized networks of
connected peers. Communication is referred as 'viral' because ideas and opinions spread like
epidemic diseases though the network via word-of-mouth. Information in social networks is
perceived as highly trusted by the users because it is based on group similarities that lead to
increase homogeneity through attitude or behavior change. The main difference between offline
and online social networks is that in the latter there are no geographical constrains and people are
able to disseminate contents to a massive audience via word-of-mouth. Social media allows
people to share almost anywhere to almost anyone “connected” on a scale that has not been seen
in the past. For example, in December 2010 a picture without make-up of the glamour rock star
Katy Perry was posted on Twitter by her husband. Even if the unflattering Katy Perry’s photo
appeared only for a few seconds, many followers retweeted the picture that was spread all over
the world and seen by millions of people in few hours. Content diffusion in social media is
independent of its source and it cannot be controlled. The likelihood to share and disseminate
information is based on the users’ choice.

Participatory
Online reputation mechanisms are directly linked to the high degrees of co-production and proactive engagement of the stakeholders in the creation and circulation of content. Indeed, the key mechanism that drives consumer empowerment is the growing willingness of users to actively engage in online conversations. People are more and more interested in writing product reviews and in sharing opinions and increasingly rely on opinions posted on social media in order to make a variety of decisions (Dellarocas, 2003). This new form of grassroots collective wisdom has been seen as a reflection of a new emerging participatory culture (Jenkins et al., 2005). “A participatory culture is a culture with relatively low barriers to artistic expression and civic engagement, strong support for creating and sharing one’s creations, and […] in which members believe their contributions matter, and feel some degree of social connection with one another” (Jenkins et al., 2005: p.3). In this scenario there is no longer a distinction between producer and consumer of culture. Indeed, Bruns (2007) introduced the notion of produsers to account for users becoming producers of digital knowledge and technology. In such a participatory media environment, the construction of corporate reputation is less subject to corporate control and intervention but it is co-created in a dynamic way together with stakeholders (Bunting & Lipski, 2000; Kozinets et al. 2010).

_Bi-directional_

Social media represents not only a _space_ for expressing opinions and ideas, but also a _fora_ where people engage in discussions in an horizontal way. Traditionally, organizations have communicated to, rather than with stakeholders (deBussy et al., 2000). By contrast, new digital technologies have enabled dialogical communication between organizations and customers and among customers. This bi-directional feature does not only provide the possibility for the
companies to reach a broad audience with low cost, but it has enabled individuals to almost costlessly make their personal thoughts and opinions accessible to the global community of Internet users (Dellarocas, 2003). The dialogical and public nature of online communication allows customers to influence public opinion by providing feedbacks on their experiences using various products and services. Gaines-Ross (2010) has coined the term “Reputation Warfare” to account for the disruptive potential of these unsatisfied and highly motivated “small-scale adversaries” (Gaines-Ross, 2010: p.70). Nowadays, several companies have experienced the reputational damage that can occur via online social media. For example, when United Airlines refused to reimburse a professional musician for breaking his guitar in 2008, he wrote a protest song and uploaded the video on YouTube. His video was seen by millions of people in few days and the news was reported by several news media. Reacting to the negative publicity, the company quickly settled a new offer (Pekka, 2010). On the other side, Dellarocas (2003) emphasized the role played by feedback mechanisms for building trust and fostering cooperation in online marketplaces, such as Amazon, characterized by large numbers of small players. This has been motivated by the fact that many traditional trust-building mechanisms, such as state-enforced contractual guarantees and repeated interaction, tend to be less effective in large-scale online environments. In online marketplaces, feedback reputation mechanisms have emerged as a viable mechanism for inducing cooperation among strangers in such settings by ensuring that the behavior of a player towards any other player becomes publicly known and may therefore affect the behavior of the entire community towards that player in the future (Resnick et al., 2000). Since online reputation and trust are the result of company’s previous behaviors towards and interactions among stakeholder’ communities, strategic reputation management discourse is
increasingly shifting on ethics and how to establish an affective relationship within online communities rather than on how to pursuing short-term interests (Pekka, 2010).

**Stakeholders’ evaluative perceptions**

Stakeholders’ perceptions of corporate attributes refer to the overall opinion about a company defined as an attitude. Attitude has been defined as “psychological tendencies that are expressed by evaluating a particular entity with some degrees of favor or disfavor […], when referring to corporate reputation we prefer to restrict the concept to the cognitive and affective responses” (Einweller, 2010: p.301). The role of subjective affective responses (feelings and emotions) is increasingly recognized as a relevant factor that drives consumer evaluations (Pham et al., 2001).

We can define affect as a valence feeling state and corporate affect as the general feelings towards the company (Aqueveque & Ravasi, 2006). Several studies have shown that people infer the direction of their preferences (liking vs. disliking) from the valence of their feelings toward the target (Schwarz, 1990) and the strength of their preferences by the level of arousal elicited by the target (Pham et al., 2001). Attitude towards companies are increasingly shaped by the opinions and feelings that circulate within digital networks. Traditionally news media were the main channel for stakeholders to gain knowledge about corporate reputation that were difficult to directly experience or observe (Einwiller et al., 2010). Today, more and more people gain knowledge about a company by searching and interpreting online signals. These signals are no longer only based on the range of comparisons between companies with similar offerings, but also on how a social network perceives the performance and quality of a company. Once people have built a picture, they share their opinions and feelings with others and “the subjective truth turns into a collective truth about what an organization is and what it should be” (Pekka, 2010:}
This is particularly true in social media platforms, such as Facebook and Twitter that are specifically designed for sharing emotions, feelings and opinions among users. Social media platforms differ from consumer reviews (such as Epinion.com) in the motivation that drive people to share. In the former the motivation is mainly emotional and refers to the need for social connections; whereas in the latter the advocacy motivation prevails (Krishnamurthy & Dou, 2008). In these networks users tend to transform messages from “persuasion oriented hype to relevant, useful, communally desirable social information that builds reputations and group relationship” (Kozitenz et al., 2010: p.83). According to Pekka (2010), social media has “the effect of presenting a collective truth” that is based on emotions and feelings of the users (Pekka, 2010: p.46). Consequently, online company reputation in social media can be seen as the general feelings and sentiments around a company then company’s achievement of a positive reputation is more and more about the ability to attract affective investments from the stakeholders (Arvidsson, forthcoming).

**Measuring online corporate reputation in social media using sentiment analysis**

Despite the fact that companies are increasingly realizing that online stakeholder voices can wield enormous influence in shaping the opinions of other stakeholders, only recently with the availability of massive data produced in social media sites and a set of new algorithms developed has it become possible to deploy data mining techniques, such as opinion mining and sentiment analysis in order to monitor and analyze stakeholder’ opinions (Pang & Lee, 2008).

Recently, several studies have shown that users opinions in digital media are better predictors of consumer choices compared to traditional indicators. Using sentiment analysis, Mishne &
Glance (2006) have shown that positive sentiment are better predictor for movies success that the volume of discussion when applied to a limited context around references to the movie in weblogs, posted prior to its release. Gruhl et al. (2005) have studied the predictive potential of online chatter in book sales. Based on an analysis of around half a million sales rank values for 2,340 books over a period of four months, and correlating postings in blogs, media, and web pages, they found that, even though sales rank motion might be difficult to predict in general, algorithmic predictors can use online postings to successfully predict spikes in sales rank. Asur & Huberman (2010) have shown that a simple model built from the rate at which tweets are created about particular topics and the sentiment expressed about a movie extracted from Twitter, can outperform market-based indicators in predicting box-office movie revenue. Jansen et al. (2009) investigated Twitter as a form of electronic word-of-mouth (eWOM) for sharing consumer opinions concerning brands. For eWOM these microblogs offer immediate sentiment and provide insight in affective toward products at critical conjunctions of the decision-making and purchasing process. The aim of the research was to investigate the possible effect of microblogging via eWOM on the brand knowledge and brand relationship. They found that nearly 20 percent contained some expression of brand sentiments. Of these, more than 50 percent were positive and 33 percent were critical of the company product. They also found that the brand sentiments for each of the 50 brands changed overtime.

Apart from the ability in predicting revenue and sales, the evaluation of users' opinions and sentiments in online social media is also a good proxy of company reputation. Indeed, sentiment analysis allows subjective perceptions, like the experience of or affective ties that consumers can construct with a company, to acquire an objective existence as observable and measurable forms
of reputation. Nowadays, more and more companies, as well as investors and marketers scan the Web, extracting reputation trends by aggregating subjective opinions (Ogneva 2010).

Furthermore, online conversations offer a constant flow of information that allows monitoring the evolution of reputation overtime. Several studies have focused on the study of topic and sentiment changes overtime in online social media. Gilbert & Karahalios (2010) have estimated information about future stock market prices based on the analysis of the emotions expressed in blogs. Based on over 20 million posts made on the site LiveJournal, they found that increases in expressions of anxiety predict downward pressure on the S&P 500 index. O’Connor et al. (2010) have found that a relatively simple sentiment detector based on Twitter data replicates consumer confidence and presidential job approval polls. They suggested that expensive and time-intensive polling can be supplemented or supplanted with the simple-to-gather text data that is generated from online social networking. However, the most interesting time-related attribute of online social media is that they allow real-time stream. An impressive application of this feature is found in the work of Sakaki et al. (2010). In their study, they used the real-time nature of Twitter for event detection based on “social sensors” (Sakaki et al. 2010). Particularly, they developed an earthquake’ reporting system that is able to automatically identify when and where earthquakes occur based on the real-time monitoring of the tweets.

Sentiment analysis

The term sentiment analysis first appears in 2001 in a paper by Das & Chen aimed to analyze market sentiment. Sentiment analysis is part of the affective computing paradigm and refers to the process of categorization of unstructured human-authored documents “based on their affective orientation, meaning the emotional attitude of the person expressing the opinion”
(Mølgaard & Szewczyk, 2010: p.1). There are three common basic approaches: full-text machine learning, linguistic analysis and lexicon-based methods. The sentimental classification in machine learning approach is based on two steps. First, a training data set is created by manually coding a set of sentences according to their sentiment. Then, an algorithm for automatically detection of the sentiment is trained according to the previous classification. The resulting algorithm is then able to detect and classify new objects (i.e. opinions) according to their sentiment polarity or valence. The linguistic analysis is inferring the sentiment valence of a text based on its grammatical structure. Linguistic analysis attempts to identify superlatives, negations, context and idioms as part of the prediction process (Thelwall et al., 2010).

The most common approach for text classification is using a lexicon. This approach requires “the creation of a knowledge base-lexicon of affective words, with additional data characterizing emotional states and relations” (Mølgaard Szewczyk 2010: p.2). In this case, we start with lists of words that are pre-coded for polarity and sometimes also for strength and uses their occurrence within texts to predict their polarity or valence (Thelwall et al., 2010). The most general sentiment classification allows the polarity classification of a text by distinguishing between positive and negative sentiment. More elaborated classifications include the identification of the strength of a sentiment. The underlying assumption is that “humans can differentiate between mild and strong emotions” (Thelwall, 2010: p.4). In this case, sentiment expressions are classified according to their valence -i.e. how positive or negative the expressed sentiment is- and arousal -i.e. level of the emotional excitation- (Hansen et al, 2010).There are several word lists labeled with emotional valence, e.g. ANEW, WordNet-Affect, OpinionFinder. The Affective Norms for English Words (ANEW) is a list of words which is considered as a reference for sentiment analysis. ANEW consists on set of verbal materials rated in terms of
pleasure, arousal, and dominance in order to create a standard for use in studies of emotion and attention. It records valence, arousal and dominance on 1034 words on a continuous scale between 1 and 9. Since this word list was developed before microblogging and it doesn’t include Internet slang, Nielsen (2011) has proposed a new word lists based on Twitter text. This Twitter-based word list has 2477 unique words and it includes 15 phrases, the score ranges from −5 (very negative) to +5 (very positive) and it doesn’t include arousal and dominance. Nielsen used a labeled database of 1000 tweets in order to compare the new list with ANEW. He found evidence that his list performs better than ANEW because of the inclusion of Internet slang and obscene words.

**Monitoring Real-time Online Corporate Reputation**

In this section, we present an open platform that is aimed to help companies, as well as researchers and practitioners to study and to monitor company reputation based on real-time stream from Twitter. Indeed, while companies and practitioners increasingly recognize the growing importance of social media as vehicles for alternative value conceptions, this kind of information is not always easily accessible for neither companies (particularly smaller and medium sized companies and non-profit organizations) nor academic researchers (like business school academics). The former might not have the necessary resources to purchases commercial research services, the latter might lack highly specialized programming and analytic skills. In addition many companies and other actors have difficulties in understanding how social media based information on alternative value conceptions can be integrated as a source of value in their operations. In the light of this, we present an open platform that can provide a common
framework for measuring, visualizing and monitoring online company reputation based on social media real-time data, and initiate a learning process around how such information can be integrated in processes of value creation.

E-Daemon is a platform aimed at exploring how stakeholders affect and sentiment in social media can be used as a proxy for integrating company’s intangible assets measurement. The platform is based on Twitter Stream API data. Twitter is a microblogging service launched in 2006 that allows users to describe their current status via short posts (i.e. tweets). A tweet is limited by 140 characters and can be posted through three methods: web form, text message, or instant message. Twitter is the most important platform for microblogging and also the social media with fast growth in the last two years. One of Twitter co-founder, Evan Williams revealed some interesting statistics regarding Twitter in 2010. In April 2010, Twitter had around 103M users; it reached 300K new users a day and 600M search queries a day.

Company reputation is “objectified” using sentiment analysis algorithm by measuring the “affective charge” of user tweets. The sentiment score of a tweet is defined as the sum of the affective words in a sentence and it is based on Nielsen affective word list (Nielsen, 2011). The resulting “affective trend” is then visualized in a real-time plot that allows the platform users to follow their target overtime. Furthermore, the platform allows monitoring the sentiment flows around a target event (e.g. the launch of a campaign or an announcement). The results are then stored in a Non-Relational Database and can be downloaded with a click in different formats. The goal of E-Daemon is to provide a tool that can be easily employed by both, business community and researchers to analyze and monitor real-time trends in social media. The innovative idea behind our platform is to use the same stream data as a common source for
several different data analyses, what we called tasks. A task is a filter that is applied to the stream-data. A user can define a specific analysis (task) and the results will be displayed in real-time. This means that the platform will be able to visualize a wide variety of different indicators of social media based reputational value that can be customized to answer to particular user needs. The platform follows an open-innovation approach. From a technical point of view, this means that the tool is developed as an expandable and programmable platform. Precisely, there are three different level of analysis in E-Daemon:

1. High level Interface, where a user creates a task via template. A template is a user-friendly pre-fielded form, where users without programming skills can specify the features of their tasks. A Template allows for a standard analysis based on real-time sentiment analysis. The user is asked to insert a target, a list of keywords, and (optional) an event or a set of events (via RSS-feed) that are displayed along with the sentiment trend in the time-line.

2. Medium level Interface, where a user is asked to choose between different algorithms, for example geo-trend visualization.

3. From the scratch, where a user can program its own filter/algorithm. This could be an interesting way to test several algorithms with the same data in real-time by creating separated tasks.

One of the great potential of this tool is the interaction among levels. Indeed, the greater the number of high-level users (let’s say “business people”) inserting keywords for targeting, the better the list. In this case, the low level users (let’s say “data miners”) will benefit of a domain specific list of keywords as identified by business people. This will allow them to improve the
quality of their algorithms that will be then used by the business people at the high-level interface. Summarizing, the open-innovation approach allows the development of two important processes:

1. *A virtuous process* of learning by doing generated by user usage. Indeed, the open approach allows an improvement of the platform *via* user generated content.

2. *Customization and plurality of algorithms*: the platform also allows customized options in order to match the needs of the companies interested in monitoring their on-line reputation.

This flexibility of the platform allows for a conceptual development of common ground for value indicators can be progressively integrated within the calculative devices visualized on the platform. We conceive of the platform as an example of what Bruno Latour has called *Dingpolitik*, a thing that has been made public and that is shaped through public deliberation to fulfill a public function in democratizing access to and development of new value indicators.

**Conclusion and Future Development**

Reputation management in online environment is becoming a central issue for companies. Corporate are increasingly loosing the control over the creation of their reputation. Indeed, corporate online reputation is more and more the result of subjective opinions shared by the users in their social networks (Pekka, 2010).
As argued by Gaines-Ross (2010), several aspects can improve the ability to manage online reputation. According to her, one of the most important elements is the high speed of corporate online actions. Nowadays most of the companies are slow moving but the ability of promptly respond to reputational changes is a necessary condition for the achievement of a good reputation (Gaines-Ross, 2010). In this paper we proposed a very proficient way for the companies to have a constant feeling of their online reputation. We presented a tool devoted to monitor and visualize online company reputation based on real-time Twitter stream. This open-platform exploits the users’ generated content for improving the company's knowledge of its reputation in social media. Furthermore, it provides a suitable framework for testing algorithms and improving the quality of domain specific data mining. Future developments will be aimed to provide a more general Online Reputation Platform that will include different sources of data, network dimension and a broader concept of reputation.

REFERENCES


