CBS COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

# Big Data Analytics in Crisis Communication & Management

**A theoretical evaluation**


**Authors:** Alexander Lundsgaard-Larsen & Mathias Gadegaard

**Supervisor:** Lasse Peter Laursen

Copenhagen Business School


Master Thesis

Cand.Merc (SOL)


**Pages:** 113 (257.024 characters)

Date of hand-in: 15-09-2016


# Abstract

This paper explore how big data analytics can contribute to the crisis communication discipline. The focus is to find the theoretical possibilities and limitations for big data analytics in crisis communication & management. Academics have done little research on this topic and it seems that no previous academic literature has made an extensive theoretical investigation about this combination. The paper approach this topic through a theoretical comparison of literature within crisis communication and the field of big data analytics. A preliminary analysis of crisis communication theory identifies five areas within crisis communication that are interesting in regards to big data analytics. The paper evaluates the use of big data techniques in each of these areas, as well as the general use of big data analytics in crisis communication. The results show that big data analytics has some potential within crisis communication & management. However, this paper also shows some clear limitations of big data analytics within the crisis communication discipline.

# Table of content

# Chapter 0 - Introduction

At this moment in time, there are clear empirical indications that the field of communication is under a current transformation. The 2016 report from the European Communication Monitor (ECM) treats some of the hottest topics among practitioners working with communication across Europe, and the report show some clear tendencies within strategic communication Topics such as big data, social media influencers (SMI's), stakeholder engagement and strategic issues related to communication channels are drawing massive attention from communication practitioners these years. Big data might be the most interesting of these topics, as 72% out of more than 21,000 communication practitioners across Europe believe that big data will change their profession in the coming years (Ansgar Zerfass et al., 2016). This number indicate that there already exists a consensus among practitioners that big data has a future in the communication field. In some areas of the communication field, big data analytics has already become an integrative part of best practice. This is observable within e.g. marketing, where both academics and practitioners agree upon the contributions of big data analytics, and successfully have integrated big data methods to solve issues and tasks (Ansgar Zerfass et al., 2016). However, the report also stresses that there is a general lack of knowledge about big data in the communication field and a lack of academic research on the topic in relation various areas within communication (Ansgar Zerfass et al., 2016). This means that it is still unclear what big data means for the future of communications (Ansgar Zerfass et al., 2016). At the current moment, only 23% of the respondents in the report believe that they have a clear understanding about what big data actually is (Ansgar Zerfass et al., 2016).

The ongoing development of online and social media means that SMI's, stakeholder engagement and strategic issues related to communication channels currently are central matters among communication practitioners (Ansgar Zerfass et al., 2016). The ECM report shows that social media is becoming increasingly important within different communication areas, whereas the importance of traditional media is on decline (Ph.d and Kforum, n.d.). 58.4 % of organizations across Europe use SMI's in their communication activities, and the focus on stakeholder engagement is moving towards online engagements such as "likes", social media comments, Facebook/Twitter posts etc. (Ansgar Zerfass et al., 2016). Online and social media is changing the discipline of communication. Currently, two of the most important strategic issues in communication management is 1) to handle the speed and volume of the information flow, and 2) to cope with the digital evolution and social media (Ph.d and Kforum, n.d.). One of the communication disciplines that is experiencing this "online revolution" is crisis communication.

## 0.1 A focus on crisis communication

The crisis communication discipline seems to be an increasingly interesting area within the field of communication. This is partly because the development of online and social media is changing the discipline, but also because organizational crises seem to occur on a regular basis. In his diagnosis of the modern society, the German sociologist Niklas Luhmann express that civilization is moving towards the "risk society" (Kneer and Nassehi, 1997, p.178). The risk society has transformed the imposition of responsibility, and the modern society now impose more responsibility on the "perceived cause" behind the accident or damage, instead of imposing the responsibility on "fate" or "destiny" (Kneer and Nassehi, 1997, p.178). The behavior in the risk society builds upon an emergent need to rationalize explanations about situations and address the responsibility of actions with negative outcomes (Kneer and Nassehi, 1997, p.178).

Modern organizations operate in an environment where they risk becoming responsible for every action they take. At the same time, organizations are becoming more aware about the benefits of a good reputation, and the efforts is takes to build it (Bracey, n.d.). A study from Harvard claims that at least 20% of the news reports in leading media must be positive if an organization wants to improve its reputation, while no more than 10% can be negative (Eccles et al., 2007). This shows that it takes more work to gain reputation, than to lose it. If an organization experience an organizational crisis it can have massive implications both internally and externally, and the reputation can take a lot of time to rebuild (Eccles et al., 2007). It is important for organizations to protect their reputation and avoid organizational crises. However, a study in the UK showed that 45% of senior executives within British companies doubt that their organization is prepared for an organizational crisis ("Crisis Communications," 2016).

Crisis communication and management is not about avoiding legal responsibility. In some cases, like the recent Volkswagen emission scandal, organizations must face the legal consequences no matter what. Nevertheless, in any case of an organizational crisis it is in the interest of the organization to come out of the situation in the best possible manner, or totally avoid the situation in the first place. Crisis communication therefore has an important role for organizations all around the world. Along with the recent tendencies in the communication field, crisis communication is a very interesting area to study.

## 0.2 Big data in crisis communication

Academics within the crisis communication discipline have developed extensive frameworks about how to deal with organizational crises. These frameworks provide a wide range of tools and methods for solving different crisis communication and management issues. In recent years, the focus has turned towards the implications and effects of the online world on crisis communication and management (Coombs, 2015, p.17). This leads to a development and integration of new perspectives about the praxis of the crisis communication discipline (Coombs, 2015, p.26). The online revolution is currently becoming an established part of crisis communication theory as due to the attention it is receiving. In contrary to this development, only few researchers and practitioners are looking into the link between big data analytics and crisis communication. A search for publications, blogs and other writings about crisis communication and big data show only little research about this topic. In the existing writings about this topic, there are indications that big data and crisis communications is an interesting match ("Datafloq - Is Big Data The Key to Crisis Management?," n.d.). However, it seems that no researchers have made a preliminary theoretical investigation about the potentials of big data analytics in crisis communication.

Big data is a new field, and academics have not yet explored the potential of big data analytics (Fisher et al., 2012, p.51). To figure out if big data has a potential in crisis communication, there is a need for a preliminary theoretical investigation. The purpose of such an investigation is to determine which areas in crisis communication that have potential for further (and more "into depth") research, and which areas it might be unnecessary to look further into. The aim of this paper is to carry out such an investigation.

## 0.3 Theoretical approach, target audience & research question

It is important to understand that the approach of this paper is purely theoretical, and the idea is to make an initial investigation that opens up for further experiments and research. The fact that this is a theoretical paper means that it has a theoretical focus on crisis communication and management, as well as on big data analytics. The paper does not carry out any practical experiments. Instead, it looks at theoretical methods in crisis communication and management, and consider where and if big data analytics has a potential to contribute. In this regard, the target audience of this paper is academics and practitioners within the crisis communication discipline (or the communication field) that want a theoretical assessment of big data analytics in crisis communication & management.

Based on this approach, the research question of this paper is the following:

*What are the theoretical possibilities and limitations of big data analytics in crisis communication & management, and how can big data analytics potentially contribute to existing methods for handling organizational crises?*

# Chapter 1 - Delimitations

Both the crisis communication discipline and the field of big data analytics are extensive in terms of academic literature. It is beyond the scope of this paper to examine the entire amount of available literature about these areas into depth. It is necessary to delimit the amount of literature within both areas in order to increase the analytical focus in this paper. The following section explain the literary choices in regards to crisis communication and big data analytics. Furthermore, this section also validates some additional choices of complementing theory.

## 1.1 Crisis communication theory

This paper understands the discipline of crisis communication through W. Timothy Coombs' book "Ongoing Crisis Communication: Planning, managing & responding". The primary purpose of the book is to offer an integrative framework for crisis communication & management (Coombs, 2015, p.1). The main reason why the paper uses this book is that the framework is coherent and seeks to include all aspects of the field. Moreover, this paper approaches the crisis phenomenon by only focusing on organizational crises. The same thing applies for Coombs' framework, which only increases its relevance. The framework is up-to-date with recent changes and developments. According to Google scholar, W. Timothy Coombs is the most cited academic within crisis communication ("Google Scholar Citations," n.d.). This means that he is one of the most prevalent names in this area and has great influence. Coombs has a considerable amount of academic knowledge about crisis communication & management, and it makes sense to draw on his experience.

Another benefit is that through the choice of this framework, the paper avoids the job of making an entire literary review of the crisis communication area, and it avoid the need to construct an extensive description of the area for analytical purposes. Instead, the paper can use Coombs' framework to depict the discipline, and based on this depiction combine big data analytics with crisis communication. Ultimately, the framework creates a structure that improve the conditions for analyzing the contributions of big data analytics, and it provide a general idea about the shape of the crisis communication area.

Despite its importance, this paper also delimitates certain parts of the framework. A general idea in Coombs framework is that crisis communication draws upon three proactive management functions (Coombs, 2015, p.31). These proactive management functions are risk, issues and reputation management. This paper delimitate threats related to risk and issues management. The reason why risk management is delimitated, is that it has more of an internal focus compared to issues and reputation management, and the sources are likewise of internal nature (Coombs, 2015, p.38). Coombs describe the risk management data derives from surveys, audits, reports about exposure etc. (Coombs, 2015, p.50). Such data is very case-specific, which makes it difficult to generalize the methods needed to identify the threats. Furthermore, the concept of big data emerged due to its large volume and how difficult it is to manage (Fisher et al., 2012, p.52). In many cases, the data in risk management does not classify as big data.

The delimitation of issues management is more difficult to make and well considered, since issues management has a partly external focus and various data sources that fits into big data analysis (Coombs, 2015, p.44+50). Issues management introduce concepts such as environmental scanning and red flags, which sets the stage for an investigation whether big data analytics is capable of detecting or analyzing warning signs linked to issues, changes, trends etc. (Coombs, 2015, p.45). However, the main reason for this delimitation is that Coombs is not specific enough in terms of explaining the technical use of these concepts. Coombs define the concepts, but fail to present any useful frameworks of environmental scanning and issue detection (Coombs, 2015, p.45). He justifies this by arguing that environmental scanning strategies are not well developed (Coombs, 2015, p.45). Hence, it is not possible to determine where exactly to apply the technologies through Coombs. Furthermore, a subsequent investigation of the area has not lead to any other frameworks specific enough to apply big data technologies in relation to environmental scanning. Big data analytics involves analytical tools, and it is important to define what they look for before they can be used (Fayyad et al., 1996, p.30). An analysis that applies big data analytics as an issues detection and environmental scanning tool is possible, but it would turn out too hypothetical and not be able to contribute with any general knowledge.

## 1.2 Big data

This paper seeks to examine if big data analytics can provide value to the crisis communication discipline. Big data analytics is the process of extracting insights from big data (Fisher et al., 2012, p.51). In order to evaluate if big data analytics is useful in the field of crisis management, this paper focus on techniques to analyze big data. By evaluating techniques, this paper identifies which kind of insights it is possible to extract using the techniques available today. Subsequently, the analysis examines if these insights are useful for a crisis practitioner in handling a reputational crisis.

Within the field of big data analytics, there are two different kind of big data techniques. These relate to either data extraction or analysis of data (Gandomi and Haider, 2015, p.139). In this paper, there is a focus on big data analytics techniques that analyze data rather than data extraction techniques. The examination of extraction techniques does not provide the reader with an overview of the possibilities of big data analytics for the crisis practitioner, which is the focus of this paper. When applying data extraction techniques, there are currently countless amount of data sources each using their own data categorizing methods and presuppositions for generating data (Tamhane and Sayyad, 2015, p.22). In this regard, the data may have many different types and shapes (Tamhane and Sayyad, 2015, p.21). This makes it very difficult to standardize the process of collecting data. The methods for data collection are often case-specific (Fayyad et al., 1996, p.33), which makes it difficult to generalize the data extraction process. As a result, it is an insuperable task to outline all possible data collection methods.

The paper put an emphasis on reputational crisis, which primarily develop based on external threats (Coombs, 2015, p.44). This makes it favorable to focus the information gathering process on external sources of information. As a result, the theoretical framework of big data highlights techniques that apply to external data. Reputational crisis mainly unfold itself in the media and on online communication channels (Coombs, 2015, p.35). Furthermore, the premise of this paper is that a crisis only exists if the stakeholders involved believes there is a crisis (Coombs, 2015, p.3). In this regard, it is a natural respond to monitor and analyze data about stakeholders in the media and on social media to acquire knowledge, so the crisis practitioners' better can act upon the situation.

This paper does not include big data techniques, which only apply to internal data. Not all organizations collect data on their stakeholders and if they do, the data might not be valuable for the crisis practitioner. In a crisis, the organization mainly collected the available data before the crisis started. Due to the new circumstances, the crisis practitioner requires fresh data to act upon (Coombs, 2015, p.118). The structures of internal data vary from organization to organization and it is difficult to generalize an analysis on this foundation. Methodically, it is difficult to make any conclusions on internal data techniques based on these circumstances.

## 1.3 Luhmann & Framing

Coombs' framework provides an overview on the crisis communication discipline but does not go into detail with all aspects of the discipline. It is directed at the crisis practitioner and guides practitioners on how to manage crisis both pre, mid and post crisis (Coombs, 2015, p.xi). According to Coombs, crises are perceptual (Coombs, 2015, p.3) and therefore, it is important to examine how the stakeholder perceptions emerge, and to which degree communication affects perceptions. Coombs provides tools to analyze communication and social phenomena, but he does not go into detail with communication, or how it relates to perceptions. Without this knowledge, it is difficult for this paper to evaluate, if it is possible to apply big data analytics. Consequently, this paper needs to draw upon additional theories to gain greater insights about the possibilities and limitations of big data analytics.

Big data analytics is a field with methods to analyze large and complex amounts of data (Gandomi and Haider, 2015, p.140). To outline the possibilities of big data analytics, it is necessary to include theories that describe the data field in more detail than Coombs' framework. In this regard, Luhmann's system theory describes the relations between the individual and society, or more specifically, perception and communication. It describes how systems functions and changes over time (Thyssen, 2012, p.694-695). Luhmann's theory on social systems provides a framework that describes the flow of communication from a social point of view (Thyssen, 2012, p.686-687). It complies with the framework that Coombs provide, as both theories takes a social constructivist approach. Knowledge on social systems provides a foundation that makes it possible to evaluate the possibilities and limitations of big data analytics theoretically.

System theory does not go into detail on how communication affects perceptions and changes the social view on phenomena. Coombs introduce the concept of crisis framing, when describing how situations turn into crises (Coombs, 2015, p.111). However, he does not go into details on the mechanisms of framing, which is necessary in order to explore the possibilities of applying big data analytics as a tool for information gathering. In extension to Coombs' framework, it is a natural choice to include additional theory on framing to examine the mechanisms in detail. Entman creates a coherent theory on framing based on a fragmented field (Entman, 1993, p.51). This makes a good basis to understand the data field from where the crisis practitioner gathers information.

The concepts of system theory and framing provide the paper with additional perspectives on how to apply big data analytics to a social context. Without additional theory on communication, the paper is not able to examine the data field and explore the theoretical implications of using big data analytics in the crisis communication discipline.

# Chapter 2 - Methodology

The following section provide insights into the methodological choices and considerations behind this paper. The first part of this section begin by explaining the early thoughts behind the paper, in addition to why the paper approach the topic theoretically. The second part of this section provide insights into the literary review of the big data field, and the third part explain the theoretical foundation of the analysis. The fourth part goes into the actual structure of the analysis, while the fifth part consider the philosophy of science in this paper.

## 2.1 - Methodical approach of the paper

Big data analytics is still a new field in academics and researchers argue that its potential has not yet been fully explored (Gandomi and Haider, 2015, p.138). For this reason, it has been difficult to find the right approach to the big data field. In early stages, there was an opportunity to collaborate with IBM, and investigate whether the supercomputer Watson has any potential as a tool in crisis communication & management. However, to examine the use of a single technique without any evidence that indicates a potential contribution is a gamble.

There is a need for an overall unification of big data analytics and crisis management & communication. Without an extensive overview of the possibilities of applying big data analytics, the crisis practitioner does not where to look for contributions of big data analytics. A "single-technology" approach could provide deep insights into the specific technique in examination, but it has never been the interest of this paper.

Another early stage dilemma was whether the study would benefit from expert interviews. Although this could be beneficial, it is difficult to find experts with a sufficient amount of knowledge in both areas. Furthermore, these interviews would not contribute to exploring the overall potential of big data analytics. Instead, it would merely describe how analysts apply big data analytics at the current state in a limited area.

The methodological choice of this paper stood between a case study and a theoretical study. However, the major issue with the case study approach is that it is difficult to draw general conclusions about the potential of big data analytics in crisis communication (Rienecker and Stray Jørgensen, 2005, p.297). Although a case study might bring some practical issues on the table, it is difficult to generalize these results. To draw general conclusions with this approach, several case studies showing the same outcome might be necessary. Moreover, a case study might not include a perspective on all the different areas of crisis communication.

After some consideration, a theoretical approach to the topic seems like the best methodological approach. By taking a theoretical approach, this paper is able to examine the combination of big data analytics and crisis communication in a broader perspective. This means that the conditions for the paper to make general theoretical conclusions improves. However, a consequence of this is that the conclusion remains theoretical, and the practical implications remain unknown.

## 2.2 - The literary review of big data

Prior to the analysis, it is necessary to get an overview of the different possibilities within the big data field. In this regard, the paper has conducted a literary review of the big data field. The results of the literary review constitute the theory part of big data analytics in this paper. A literary review provides a general understanding of a field as well as a theoretical foundation for further research (Ridley, 2008, p.30). This paper uses the following definition of a literary review:

”*a systematic and thorough search of all types of published literature in order to identify as many items as possible that are relevant to a particular topic.”* (Ridley, 2008, p.29).

In this paper, the literary review is important for multiple reasons. An important part of this paper is to improve the understanding of the big data concept. In this regard, the literary review contributes through the identification and establishment of a suitable big data definition. Finding the right definition of big data is important, because it defines the big data techniques that this paper search for. The literary review is instrumental for the discovering of these techniques, and it is essential when understanding the processes of big data analytics and the techniques in a broader perspective.

The main databases used in the literary reviewing process has been Google Scholar and the CBS library database. A search in these databases found a large number of articles about different big data techniques. However, the focus of the literary review has been to find articles that outline the big data field and categorize the currently existing techniques. This procedure provided an elaborated overview of the current technological possibilities. One thing this process brought into light is that the articles primarily categorize after types of data. Different data require different techniques for processing.

The overview of the big data field made it easier to conduct a systematic examination of the literature. In this regard, this study has carried out an extensive research on the various big data analytics techniques and methods, in order to get a comprehensive understanding of the field. The idea behind the literary review is to create a comprehensive toolbox of the current big data techniques, as well as an overview of the current technological possibilities. This is important for the analysis as it provides the necessary knowledge to analyze whether big data analytics can contribute in the crisis communication discipline.

Numerous variations of techniques exist to cope with the differences in data, and every technique adapts to the individual case. Thus, it is an exhaustive task to examine the extensiveness of every variation. As an alternative, the literary review has scanned the field for articles that apply big data techniques, in order to identify the overall techniques used. The paper categorizes the techniques after which type of data they are able to process.

## 2.3 - The foundation of the analysis

This paper seeks to analyze the application of big data techniques in crisis communication in order to gain new insights about the combination of big data and crisis communication. Thus, expanding the tools available in crisis management and set the stage for the crisis practitioner to implement big data techniques in the crisis communication discipline.

The study in this paper evolves around the theoretical framework of W. Timothy Coombs, which constitutes the paper's understanding of crisis management. The framework provides a picture of the general practice for crisis management. This is the Archimedean point, which the rest of the study builds upon. The framework of Coombs will thus be the paper's field of inquiry and replace any need for empirical data. In this regard, the framework functions in the same manner as empirical data as it is the starting point of the study, which the analysis will build upon.

The task of examining the entirety of Coombs' framework is too comprehensive for this paper. Therefore, the paper conducted a preliminary analysis of Coombs' framework to select the most interesting areas to explore. In this regard, the paper identified five areas, which involves information gathering. These are expectation gaps, stakeholder salience, situation awareness, social media monitoring during a crisis and crisis evaluation. The reasoning behind this choice is that big data analytics involves methods to extract insights from large amounts of data. Therefore, it is natural to focus on areas in Coombs' frameworks, which involves information gathering. This is where big data analytics has something in terms of contributions. The analysis goes into depths with each of these five areas to examine the potential use of big data analytics.

## 2.4 - Analytical structure

### 2.4.1 - Explaining the structural design

The analysis consists of three parts with each their own purpose. In combination, these parts aim to uncover the potential contributions of big data analytic within the crisis communication discipline. In the first part, the analysis attempts to explore if it is possible to apply big data analytics methodologically to the crisis communication discipline. Furthermore, the section outlines the possibilities and limitations of applying big data analytics to communication. Before the analysis more specifically can evaluate where big data analytics has any contributions, this is a necessary step.

The second part of the analysis seeks to explore the potential contributions of big data analytics within five areas in Coombs' framework that require information gathering. This part seeks to guide the crisis practitioner to areas where big data analytics might contribute. The analysis explains the mechanisms of how to apply big data analytics to the field of communication in detail.

The last part of the analysis creates an overview of all big data techniques and outlines their potential contributions to the crisis communication discipline on a general level. It is too comprehensive for this paper to discuss every technique in each of the five areas in the second part of the analysis. Therefore, the last part is necessary to discuss the potential of each technique on a more general level. Simultaneously, this part ensures to consider all techniques towards the five areas in Coombs' framework.

### 2.4.2 - Methodical approach in the first part of the analysis

The first part seeks to examine the transition from communication to data. This section draws upon Luhmann's social system theory and the concept of framing to analyze the communication in the perspective of crisis management. These theories provide a deeper understanding of a crisis in a social context. Using this understanding, the paper can better assess how to apply big data techniques methodically. The concepts offer an understanding of the field of communication and perception and help to outline the possibilities and limitation for using communication as data. Thereby examining what sort of data is quantifiable and works as data for big data analytics. This helps the analysis to determine which techniques that are applicable for analyzing communication and have the potential to assist the crisis practitioner. The first part of the analysis creates a foundation for the subsequent parts by highlighting potential and limitations of applying big data techniques.

### 2.4.3 - Methodical approach in the second part of the analysis

The second part of the analysis builds upon a preliminary examination of Coombs' framework. A thorough examination of Coombs' framework is too comprehensive for this paper. In this regard, the examination identified five stages in Coombs' framework that concerns information gathering and analysis. The given reason is that big data analytics is a method to gain insights through analysis, and these stages are where big data analytics has the most potential to contribute. In each of the stages, the analysis outlines what kind of information the crisis practitioner requires according to Coombs' framework. On this basis and with the help of insights from Luhmann and Entman, the paper evaluates which big data techniques that have potential to assist the crisis practitioner. As the study does not involve any empirical exploration of the use of big data, it is not possible to identify practical implication. By comparing studies from related fields, where similar techniques apply, this study is able to give a rough estimation on whether big data analytics provides value in this area. The purpose of this approach is to evaluate and locate any potential to use big data analytics.

The techniques that has the potential to assist the crisis practitioner form part of a comparative analysis with the traditional methods that Coombs suggests at this stage. To determine the value of big data analytics, it is necessary to do a comparative analysis with the tools that Coombs provide in his framework. This part of the analysis provides a general outline of the advantages and disadvantages of using big data analytics. The analysis determines whether big data analytics is able to replace or complement the existing information gathering methods. As a result, the paper estimates the value of using big data analytics to gather information for reputational crises.

### 2.4.4 - Methodical approach in the last part of the analysis

The last part of the analysis takes a closer look on all the big data techniques and evaluates whether or not they have any potential within the discipline of crisis communication. This evaluation builds upon studies from the big data analytics field and the ontological circumstances when analyzing communication. As a result, the paper is able to rule out certain techniques and recommend other techniques for further empirical studies.

## 2.5 - Philosophy of science

According to Coombs, a crisis is always perceptual (Coombs, 2015, p.3), which is one of several social constructivist features found in the framework of Coombs. He also describes how media and other stakeholders frame situations into crises (Coombs, 2015, p.110). These are social constructivist characteristics. As the theoretical framework of Coombs is the foundation of this paper, it is natural for this study to use the same social constructivist approach. Coombs consider a crisis be to a social constructed phenomenon, which affects the way the crisis practitioner examines and gather information about the crisis. From this perspective, it is possible to obtain great knowledge of the crisis by collecting the conceptions from all the stakeholders.

A premise in this paper is that knowledge regarding human interactions is socially constructed and phenomena are contingent and thus not universally applicable. According to social constructivists, certain phenomena is in fact only generated and maintained through various social practices (Hviid Jacobsen et al., 2011, p. 230). These phenomena is not able to exist independently of the social activity they are embedded in (Hviid Jacobsen et al., 2011, p. 230). In this regard, the establishment of the social reality does not depend on an actual physical object.

An actor's intention, beliefs and current applicable knowledge constitutes the external behavior and the sum of social interactions constitutes the social reality (Hviid Jacobsen et al., 2011, p. 244). Phenomena are both historically changeable and vary from society to society. As a result, it is difficult to establish universal laws regarding these phenomena. This idea is compatible with the ontological perspective in Luhmann's system theory. According to Luhmann, the environment outside systems is too complex for the human mind to comprehend (Kneer and Nassehi, 1997, p.44). As a result, the social and psychological systems seek to reduce the complexity in order to create meaning in the environment outside systems (Thyssen, 2012, p.696).

The ontological perspective in this paper provides a dynamic picture of how to apply big data techniques, and what kind of information, the techniques can find. For example, when predictive analytics finds interdependency between two variables, many analysts consider the correlation universally applicable. Analysts use correlations between variables to predict the future and to analyze societies (Gandomi and Haider, 2015, p. 143). From a social constructivist perspective, variables that in some cases are interdependent may have no correlation in other cases. When analyzing social interaction, variables may vary over time and space and the social constructivist perspective provides a more dynamic conception of the variables (Hviid Jacobsen et al., 2011, p. 230).

# Chapter 3 - Theory

## 3.1 - Big Data Theory – the literary review

In the 21st century, there has been an exponential increase in data, which represent the daily activities in society. This increase is particular due to launching of the social media and smartphones and there is no indication that this development does not continue (Fisher et al., 2012, p. 50). Of all the data on the internet, 90% have been created in the last two years (Tamhane and Sayyad, 2015, p. 18). The development have created the concept “big data”, which is relatively new in the field of research and was not widespread in the academia before 2011 (Gandomi and Haider, 2015, p. 138).

Today, big data is in particular used within the fields of marketing, security, business optimizing, public health, science and financial trading (Jagadish, 2015, p. 49). Hence, the use of big data is drawing the attention from every segment in society. The use of big data is increasing tremendously and society is yet to figure out the full extent of its potential.

It is often required to be innovative when using big data analytics, as the possibilities are not always straightforward (Gartner Inc., 2016). For example, the used of social media data from Twitter has been used to predict the sales of iPhones. The analysts in the study were able to develop a social model by analyzing social media activity and text sentiments, which have a very strong correlation with iPhone sales. (Lassen et al., 2014)

Additionally, it is possible to measure a country’s economic growth from outer space. By analyzing satellite photos taken at nighttime, the analysts found a strong correlation between the volume of lights at night and a country’s Gross Domestic Product (GDP). The GDP is a good indicator in measuring economic growth. (Henderson et al., 2012, p. 994)

The following section attempts to outline a definition of big data as well as introduce some of the major challenges when processing and analyzing big data. The HACE theorem is there to introduce a big data processing model for data mining. Furthermore, the KDD process model functions as a general model for data analysis. Lastly, the paper creates a framework of the different techniques that exist in the theoretical field of big data analytics. This creates a toolbox of techniques for the paper’s analysis.

### 3.1.1 - Definition and challenges of Big Data

The definition of big data has historically changed and there is a fine line between regular data points and big data (Fisher et al., 2012, p. 51). As the name implies, big data includes a large quantity of data but the definition of large is relative to the technological advancement. The size, that analysts consider big, has grown in accordance to Moore's Law, which states that the overall processing power of computers will double in two years (Fisher et al., 2012, p. 52). The development of better hardware and software, as well as big data techniques, makes it easier to handle and process large sums of data (Fisher et al., 2012, p. 52). According to Fisher et al., "*Big Data means data that cannot be handled and processed in a straightforward manner*" (Fisher et al., 2012, p. 52). Thus, the definition of big data relates to how difficult it is to process and not only its actual size. During the conduction of this literature review, this paper repeatedly encountered two different definitions of big data analytics. The definitions use the foundation of respectively the three Vs and the HACE-theorem, which use two different approaches to describe characteristics and challenges to big data. The three Vs describe the challenges with handling and processing big data, while the HACE-theorem focus on the data sources and the mining of big data. However, both definitions describe big data as a new phenomenon and in relation to the new challenges that comes with the enormous amounts of data available. This paper focus on the analysis of big data, rather than the data mining process. Therefore, the first definition is more suitable and apply for this paper.

### 3.1.2 - The Six Vs

Laney (2001) introduce three dimensions of a data set in order to highlight some of the big challenges when processing big data. These are: volume, velocity and variety (Laney, 2001, p.1-2). The dimensions describe how difficult the data is to handle and process (Laney, 2001, p.1), which makes them a good indication to distinguish big data from regular data. Hence, the three dimensions has later become a commonly used framework to describe and define big data (Chen et al., 2012, p. 1182).

Gartner Inc. defines big data in their online IT glossary in these terms:

"*Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."* (Gartner Inc., 2016).

- **Volume:** Data volume describes the amount of data collected within a specific field. Organizations see data as a tangible asset and are more reluctant to discard it as it might turn out to be valuable. Therefore, organizations often collect and store enormous amounts of data without knowing how to create use of it. Data comes in various formats and the volume of video data takes up more space than numbers, which is why the volume is always relative to the datatype. (Laney, 2001, p. 1)

To overcome the challenge of limited storage, it is popular to analyze the data as computers produce it. Thereby it is not necessary to storage the data for future analysis. This method is called real-time analytics and is used progressively more, as the volume of data increase. The downside of real-time analytics is that it is not possible to go back and change the techniques or correct potential errors. (Gandomi and Haider, 2015, p. 138)

- **Velocity:** The velocity of data refers to the speed in which computers create the data but also how crucial it is to analyze and act upon the data. Velocity is an indication of the workload of the analyst. The amount of data generated is rapidly increasing, which increases the velocity of data. (Laney, 2001, p. 2)
- **Variety:** The level of Variety in the data refers to the structural heterogeneity of the data (Laney, 2001, p. 2). Only approximate 5% of the data collected is in the form of structured data and the vast majority is either semi-structured or unstructured (Gandomi and Haider, 2015, p. 138). Structured data refers to data, which is found in a spreadsheet or in relational databases (Gandomi and Haider, 2015, p. 138). This kind of data is typically gathered and sorted by machines and is used in predictive analysis ("Big Data Analysis Techniques | Slice and Dice," n.d.).

  Audio, video, text and images are examples of unstructured data, which machines are not sufficient to analyze due to the lack of structure ("Big Data Analysis Techniques | Slice and Dice," n.d.). This type of data is generated by humans, which makes it unpredictable, complex and difficult to manage ("Big Data Analysis Techniques | Slice and Dice," n.d.). The definition of semi-structured data is not rigorous but often refers to data with user-defined data tags that makes them readable for computers. (Gandomi and Haider, 2015, p. 138)

IBM, SAS and Oracle have later introduced four additional dimensions to describe challenges in dataset. These are Veracity, Variability, Complexity and Value. (Gandomi and Haider, 2015, p. 139)

- **Veracity:** Veracity refers to the unreliability of the data, which does not relate to the method of data mining. For example, stakeholder sentiment on the social media has a low veracity, as it entails human judgement. The processing of data with low veracity involves much uncertainty, which affects the validity of the results. (Gandomi and Haider, 2015, p. 139)

- **Variability and Complexity:** The velocity of the data is often inconsistent and comes in different flow rates. Variability describes how much the data flow rates can vary over time. Complexity refers to the challenge of connecting, cleansing and matching data from different sources. (Gandomi and Haider, 2015, p. 139)
- **Value:** Big data often has a low value density. This means that each data point has a low value relative to the volume of the dataset. If a data set has a low value, it can be difficult to make any reliable analysis. To create valuable outcomes, it requires large amount of data because each data point has a low individual value. (Gandomi and Haider, 2015, p. 139)

### 3.1.3 - HACE theorem

Compared to the six Vs, the HACE Theorem focus on the practical implications of data mining and processing from multiple autonomous sources. By emphasizing on some of the characteristics of big data, it results in the following definition:

*"Big Data starts with large-volume, heterogeneous, autonomous sources with distributed and decentralized control, and seeks to explore complex and evolving relationships among data."*(Vilas, 2013, p. 12).
The paper does not apply this definition of big data, when conducting the literature review. Instead, the paper used the previous definition of the 3 V's.

Big data analytics often involves data collection from various sources, which creates heterogeneous data. This occurs since the data sources have their individual techniques of data collection and have their own way to best present their own data. Hence, it can be very difficult to combine and correlate data from different sources, if the data is incomparable. Each data source develops their own autonomous data, as the sources do not have a centralized control to connect them. As independent sources, they are not subject to any control outside of their own system. (Vilas, 2013, p. 12)

By "complex and evolving relationships", the definition refers to the continuous evolving and interrelated associations between individuals in society. It is rather difficult to capture the essence of associations through a set of data. Often these associations are lost when analysts create and categorize the data. When an information system is reducing social activity to data, it will inherently treat each individual as an independent entity and rarely consider their social connections. Furthermore, by analyzing limiting data and variables it is impossible to see the full picture, which may result in shortcomings. (Tamhane and Sayyad, 2015, p. 19).

A big data processing framework has been build based on the HACE theorem. The following figure shows three different tiers in a model, which describes the challenges, when processing big data. It is important to read the model from the inside out in order to understand it, since the first tier is in the middle and the third tier is at the outer ring. The three tiers describe the three different levels of data accessing and computing, where tier number one is the most basic. (Vilas, 2013, p. 13)

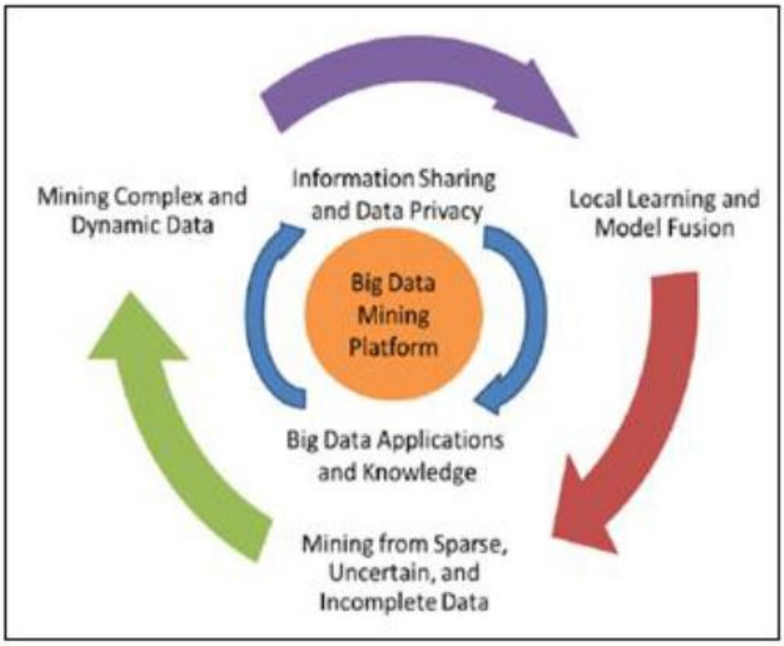


Figure 1 shows a Big Data processing framework (Vilas, 2013, p. 13)

- **Tier 1:** The first tier concerns the hardware that is necessary when processing big data. Due to the enormous volume of data, which the servers need to store and analyze, it require large amount of storage space and computer processing power. One computer is rarely enough to manage the demand in big data analytics. Therefore, multiple storage spaces and processors are required, which make things more complicated and costly. (Tamhane and Sayyad, 2015, p. 19)
- **Tier 2:** Data often involves sensitive data such as medical records and banking transactions, which is too private to be available for everyone. The examiner often has limited access to the data or sometimes the data source removes several of the data points. The reason for this is so an individual person cannot be identified through the data. Furthermore, tier two also include domain and application knowledge. When designing algorithms and modelling of data, it requires domain specific knowledge. Otherwise, it is difficult to understand and analyze the incoming data. (Vilas, 2013, p. 13)
- **Tier 3:** There are two approaches to select from, when combining data from several sources. The first is to have each source make their own statistical conclusions based on their own data and then compare the results from the various sources. The benefit of this approach is that the source has greater local knowledge of the data. However, the sources might not see the greater picture and it may lead to biased conclusions and models. The other approach is to collect all

the data at one centralized database, and then make the analysis. Although, this creates a challenge of fusing and organizing the data so it is ready for analysis. (Tamhane and Sayyad, 2015, p. 19)

Big data will often involve meager, tentative and incomplete data. If the date is meager, then there are too few data points to make any decisive conclusions. With tentative data, the data is subject to inaccurate distributions. This is due to some sort of disturbance. Incomplete data is often caused by a defect sensor node or some regular policies to intentional skip some values. However, most modern data mining algorithms have built-in solutions to handle the absent of data. The last part of the third tier involves the complexity of the data. Complexity has many different representations including semantic relations in data, complex association networks and diverse data types. (Tamhane and Sayyad, 2015, p. 19)

### 3.1.4 - The KDD process model

Many different methodological approaches exist within the field of data analytics. Often there is a need to design and adjust the methods for the individual project. Therefore, it is difficult to outline a detailed formula for data analysis in general. The KDD process model aims to create a framework for the variety of activities in data analysis and how they fit together. (Fayyad et al., 1996, p. 29)

The KDD process is defined as:” *The nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.”* (Fayyad et al., 1996, p. 30). Data mining is the

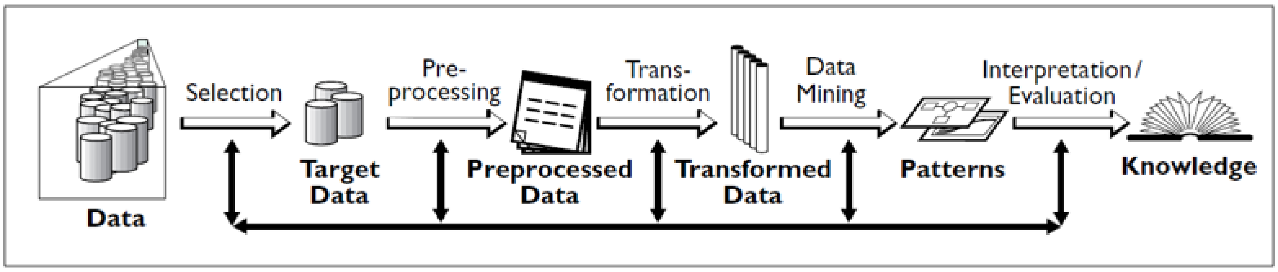


Figure 2 shows the KDD process model (Fayyad et al., 1996, p. 29).

activity of fitting models or determining patterns from data in order to create inferred knowledge. The process of determining which models produce useful and accurate knowledge will however, always require human judgement. Thus, it will not be possible to streamline the entire methodological process of data analytics. (Fayyad et al., 1996, p. 29)

The KDD process model follows these nine steps:

1. First, it is important to develop an understanding of the application domain and gather all prior knowledge, which is essential for the rest of the process. After this, analysts make the objectives for the data analysis. There is a good chance that the analysts readjusts the project later in the process, if the foundation is poor. (Fayyad et al., 1996, p. 30)
2. The initial knowledge foundation obtained in step one determines the selection of data. Based on availability, the analysts select and fuse the data together into one data set. Sometimes it is necessary to make compromises in the data selection, as not all data is available. If some important attributes are missing, the study might fail. Therefore, it is better to have too much data, than too little. (Fayyad et al., 1996, p. 30)
3. The model enhances data reliability in this step and it involves pre-processing and cleansing of data. In this process, the analysts remove noise and outliers as well as dealing with missing values. If a variable has missing data, a statistical prediction model can predict and insert data. (Fayyad et al., 1996, p. 30)
4. In this stage, the analysts reduce and project the data that involves dimension reduction and attribute transformation. The former effectively reduce the number of variable under consideration, while the latter, and refers to the discretization of numerical attributes and functional transformations. The analysts need a good understanding of the data and the project in this step, as the transformation of data is very projects specific. (Fayyad et al., 1996, p. 30)
5. There are two different major goals in data analysis: prediction and description. At this stage, it is time to decide the purpose of the model, and what type of data mining to use. These could be for example regression, clustering or classification. (Fayyad et al., 1996, p. 30)
6. This step involves choosing the data mining algorithms and methods for searching for patterns in the data. It is essential that the method reflect the objectives of the project. For instance, sometimes understandability is preferable over precision, if neutral users should understand the model. (Fayyad et al., 1996, p. 30)
7. In this step, the analysts apply the mining algorithms and they put them into practice. The algorithms might need adjustments until they yield a satisfying result. (Fayyad et al., 1996, p. 30)
8. When the modelling is complete, it is time to make an interpretation of the results and consider what affect the pre-processing steps might have caused. It is important to present and visualize the results in an understandable way for the end user. (Fayyad et al., 1996, p. 30)

9. Finally, the end user incorporates the discovered knowledge into the intended settings. This step involves many problems, as the analysts discovered the results in a static snapshot, while the end user incorporate the results into a dynamic setting. (Fayyad et al., 1996, p. 30)

## 3.1.5 - Big Data techniques

Due to the volume and variety of big data, it is difficult to gain an overview of the data and identify trends. To process the data, there is a need for different tools and approaches, as no single approach can create a full picture. Therefore, the analyst is required to adjust the approach after the objectives of the analysis. (Fisher et al., 2012, p. 50)

The following part of the paper briefly sketches an overview of the applicable techniques within big data analytics. This section categorizes the techniques after what the kind of data the techniques can process and analyze. Big data analytics is able to analyze data in the form of numbers, text, audio, video and social media relations (Gandomi and Haider, 2015, p. 140). Given the extensiveness of the techniques, an exhaustive list of techniques would be beyond the scope of this paper. Accordingly, the aim of the following sections is to give the reader a brief insight into the various techniques within big data analytics.

### 3.1.5.1 - Machine learning

Due to the complex nature of big data, it is very difficult to comprehend. Machine learning are to solve this challenge, as machines are able to analyze hundreds of variable simultaneously. Analysts use it for any kind of analysis and both structured and unstructured data. Machines are able to learn by analyzing data through trial and error and are able to learn without explicitly programming. Normally, the computers learn by analyzing a database where humans already predefine the results. This way, it determines whether the result is correct and which method that is superior. When the machine is sufficient in analyzing the data, it is ready to analyze raw data. If humans conducted the same algorithms, it would be an extensive task. (Digital Reasoning, 2014, p.3)

By using machines to identify patterns and correlations, it is possible to discover connections that might otherwise stay unnoticed. Machines are able to use an inductive approach on data analysis that does not include presumptions or knowledge about past relations. Because of the nature of big data, it can be difficult to manage. Machine learning create models and identify patterns that would be very time consuming for scientists and analysts to generate. (Intel IT Center, 2013, p.2)

### 3.1.5.2 - Techniques for structured data

Big data analytics for structured data use techniques that already exist within statistics. There are three different groups of techniques for structured data. These are descriptive analysis, predictive analysis and prescriptive analysis. Descriptive analytics collects data about the past or at a certain event. The technique analyzes and visualizes the data to show what happened in the past. It is possible to do descriptive analytics in real-time to understand and identify any irregularities. (Ernst & Young, 2014, p. 6)

Predictive analytics includes a variety of techniques that predict future outcomes based on the current data available. The techniques are valued in almost every field and are widely used within research. In general, predictive analytics seeks to find correlations and uncover patterns in data. Some techniques, such as moving averages, will try to discover historical patterns in the outcome variable and extend these patterns to predict future outcomes. Other techniques, such as linear regression models, will attempt to find correlations between the dependent and independent variables in order to make predictions. (Gandomi and Haider, 2015, p. 143)

Prescriptive analytics evaluates different scenarios and alternatives to determine the best possible option. It uses already established algorithms and rules to forecast the future outcomes of the available choices. Thereby, organizations to determine which strategy to select. (Intel IT Center, 2013, p. 2)

When applying statistical methods to big data, there are a few deviations from the conventional statistical methods. For instance, when analyzing smaller samples from the population, there is a chance that the correlations are a pure coincidence. This is not as relevant, when working with big data, because the technique includes a far bigger part of the population. Therefore, the statistical significant levels are mostly irrelevant in big data. Furthermore, many conventional methods on smaller samples require too much processing power on a large scale. Consequently, there is a need for new methods in order to capture the full potential of predictive analytics in big data. (Gandomi and Haider, 2015, p. 143)

### 3.1.5.3 - Text analytics

Text analytics refers to techniques that extract information from textual data. It is rather difficult to extract usable information from text, as it most often come in the form of either unstructured or semi-structured data.

#### 3.1.5.3.1 - Information extraction

Information extraction is a technique, which analysts widely use within the fields of biomedical research and finance and can extracts basic structured information from unstructured text. For example, given the following sentence: "*In 1998, Larry Page and Sergey Brin founded Google Inc.",* it is possible to extract the following information: Founder Of (*Larry Page*, *Google Inc.*), Founder Of (*Sergey Brin*, *Google Inc.*), Founded In (*Google Inc.*, *1998* ). (Aggarwal and Zhai, 2012, p. 11)

This technique provides value within the field of biomedical research to look through documents in order to find new scientific discoveries on particular genes. Furthermore, financial experts use the technique to identify all company takeovers that have taken place over a specific timespan. Information always have a context and therefore, a computer cannot simply search through documents looking for specific names and words. For example, a computer can either understand "JFK" as the person John F. Kennedy but also the location JFK International Airport. Hence, it is necessary to create algorithms that analyze and understand the context of the information. (Aggarwal and Zhai, 2012, p. 15)

#### 3.1.5.3.2 - Text summarization

The primary use of text summarization is to subtract and convey the key information in a text. Financial experts use this technique to summarize stock market updates. Furthermore, the social media forum Reddit.com applies it to summarize the content of posted articles for the users. In a study conducted by the US government, they found the summarization approach very useful in determining whether an article was relevant to read, by discarding 77% to 90% of the text. Thereby, the technique is very useful as a tool to remove all irrelevant articles. (Hahn and Mani, 2000, p. 34)

Broadly speaking, there are two different approaches for text summarization: the extractive approach and the abstractive approach. The extractive approach divides the text into sentences and then weight the sentences by matching phrasal patterns or by analyzing the sentences' lexical and statistical relevance. Thereby, the technique lists the most important sentences in a chronological order although there will most likely not be any coherence in the summary. This approach is simpler and require less processing power, as it will not require any understanding of the text. (Hahn and Mani, 2000, p. 30)

In contrast, the abstractive summarization technique involves extracting semantic information from the text. This approach requires using heavy processing power due to natural language processing (NLP) algorithms. The summaries contain text units that is not present in the original text, which means that the computer will need to process the text in order to summarize it. As a result, the abstractive summarization technique tends to generate summaries that are more coherent. However, the approach is far more complicated and require much more computer processing power. (Hahn and Mani, 2000, p. 32)

#### 3.1.5.3.3 - Question answering

It frequently occurs that experts in other fields than IT require knowledge that is only retrievable through big data analytics. For this purpose, the question answering technique assist people, which do not have the skills or tools to make their own analysis. The technique functions through a platform that uses machine learning and NLP to reveal insights from big amount of unstructured data. Furthermore, it continues to learn and improve through interaction. For example, if asked, "What is the treatment for this disease?" it will search through all research documents on the topic and list all possible treatments of the disease. (Song et al., 2015, p. 102-103)

Two examples of commercial question answering systems are Apple's Siri and IBM's Watson. These systems are used and have been implemented within marketing, finance, healthcare and education (Gandomi and Haider, 2015, p. 140).

The question answering technique is dependent on advances natural language processing (NPL) just as the abstractive summarization technique. The process requires three sub-components. First, NPL analyze the question, in order to understand what the user ask of the system. Second, the technique finds relevant documents either in a closed database or through a search engine. Lastly, the computer analyzes the documents and ranks them after how well they answer the question. To answer the question, the highest ranked candidate applies. (Song et al., 2015, p. 103)

#### 3.1.5.3.4 - Sentiment analysis

Sentiment analysis is a technique that mines people's opinions on entities such as organizations, product, individuals and events. This is thoroughly done within the fields of marketing, finance and the political and social sciences (Gandomi and Haider, 2015, p. 140). In the past, sentiment analysis have been used to predict the outcome of the American presidential election by analyzing Twitter messages (Colleoni et al., 2014, p. 317).

Another use of this technique is to calculate an organization's reputation, where the more traditional method would be through statistical surveys. While using sentiment analysis might not be the most accurate way to calculate an organization's reputation, it still has several advantages. It requires less resources to execute and can be done relatively quickly, at any time of the day. Furthermore, it is possible to get a live calculation on the reputation from day to day. However, it is not possible to examine different demographic groups, as you are limited to the users of the social media. (Colleoni et al., 2013, p. 319)

There are three different approaches used within sentiment analysis. These are: document analysis, sentence analysis and aspect analysis (Liu, 2012, p. 6).

Document analysis attempts to determine whether a whole document express a positive or negative sentiment by looking at the adjectives and verbs, which are positive and negative. This method is not completely accurate in determining every sentiment separately. However, by analyzing a whole document as one sentiment, the errors will have no significance on the outcome of the sentiment. The weakness of this method is that, it is using the assumption that the whole document expresses a sentiment towards a single entity. If the document evaluates more than one entity, the method cannot distinguish between different entities. (Liu, 2012, p. 30)

In contrast, sentence analysis operates by dividing the whole document into single sentences. Before, the sentences can be analyzed, it is necessary to distinguish between subjective sentences and objectives ones (Liu, 2012, p. 49). Therefore, this technique tends to be more complex and require more resources. (Liu, 2012, p. 49) Socher et al. have recently developed a method to analyze the sentiment of a sentence by an 85% accuracy, which is 5% more accurate than some of the best tools on the market (Socher et al., 2013, p. 1).

By applying the aspect analysis technique, it is possible to get a more nuanced picture of the stakeholder' sentiment. The technique does this by examining the document for all sentiments and then identifying which aspect of the entity, the sentiment refers to. For example, a smartphone has different features, which each might have their own sentiment. Thereby, it is easier for the manufacturers to collect information on what the stakeholders like and dislike about the product. (Liu, 2012, p. 59)

### 3.1.5.4 - Audio Analytics

Healthcare and customer call centers are the primary fields where audio analytics has gained ground (Gandomi and Haider, 2015, p. 141). Audio analytics can help gain insight into customer behavior, improve costumer experience, evaluate agent performance and monitor compliance with different policies (Gandomi and Haider, 2015, p. 141). By applying real-time analytics, audio analytics systems can give selling recommendations based on past and present interactions and provide feedback to callers during their conversation (Gandomi and Haider, 2015, p. 141). In healthcare, audio analytics can provide aid to diagnose and treat patients that have certain medical condition that affect the patient's communication patterns such as schizophrenia, cancer and depression (Neustein, 2010, p. 305).

Audio analytics have two primary techniques to analyze speech in audio, which and transcript-based approach and phonetic-based approach. The transcript-based approach, which also has the name large-vocabulary continuous speech recognition (LVCSR), involves a two-step process. First, the technique transcribes the spoken language using automatic speech recognition algorithms that match sounds to words. It identifies words based on predefined dictionaries. Thereafter, it is possible to apply various text analytics techniques. (Krishnan et al., 2014, p. 507)

In contrast to LVSCR, the phonetic-based approach works with the phonemes, instead of analyzing text. A word forms when a communicator puts multiple phonemes together. Phonemes are the smallest units of speech in a language. For example, the articulation of "e" have multiple pronunciations but with the help of phonemes, we can separate them from each other. Within a language, there are only a few tens of unique phonemes. After the phonemes are in text form, they are ready for a computer to process and analyze the data for further output. (Krishnan et al., 2014, p. 508)

### 3.1.5.5 - Video Analytics

The big data techniques in video analytics is still in its infancy compared to other types of data analytics. Furthermore, there are not as many fields that have adopted the techniques of video analytics to the same extent as other big data analytics techniques (Gandomi and Haider, 2015, p. 141). One of the massive challenges in video analytics is the volume of the data. In comparison, one second of high- definition video is equivalent to 2000 pages of text in size. Furthermore, 100 hours of video is uploaded to YouTube every minute ("YouTube Statistics," 2016). The increase in surveillance cameras and video-sharing websites creates a need for better video analytics technologies and help push the development forward (Regazzoni et al., 2010, p. 16).

One of the big area of video analytics, where there is a great need for development is in the field of automatic video indexing. The emergence of widespread online videos creates a need for indexing and categorization of videos. Though it is not yet possible to index videos on its visual content, techniques are able to index videos based on soundtracks, transcripts and metadata. (Gandomi and Haider, 2015, p. 141)

In retail, video analytics collects demographic information about customers, such as ethnicity, gender and age. Furthermore, it is possible to measure costumers dwell time at different shelves, detect movement patterns and monitor queues. This kind of information can be used for product placement, cross-selling, promotion design and staffing. There is a great potential of studying buying behavior of groups, since there will often be lost information when they only count for one person at the cash register. (Gandomi and Haider, 2015, p. 141)

The main use of video analytics in recent years has been within the field of surveillance systems. Replacing labor-based surveillance systems with automatic systems is cost-efficient and more effective. According to a study, security personnel cannot remain focused on a surveillance task for more than 20 minutes, which would make automated surveillance superior. (Shan et al., 2012, p. 309) Surveillance systems deployed today are not yet capable of analyzing complex actions. This is a big deficiency, since the millions of surveillance cameras deployed today cannot analyze the data in real time and provide aid with terrorism, crime and accidents. Instead, the data is stored on a server for post-event video forensics at best. (Regazzoni et al., 2010, p. 16)

In the field of surveillance system, which is the fastest developing field in video analytics, there are three main direction concentrations. The first direction focuses on visual event modeling and algorithmic studies to consistently classify and detect visual events and anomalies. The second direction is on the cooperation of several cameras rather than the interpretation of the data from a single camera. By creating a system that can fuse the data of several cameras on the same visual phenomenon, it will improve the overall surveillance performance severely. The last direction is in real time analytics, where the technicians implement the algorithms directly in the camera to avoid the challenge of volume. As of now, this technique is very expensive and the camera do not have the required processing power to make complex algorithms. (Regazzoni et al., 2010, p. 17)

### 3.1.5.6 - Social Media Analytics

In recent years, there has been a radical increase in social media channels, which creates structured and unstructured data based on the social interactions of the users. There exist many different social media platforms, which fall into the following categories: social networks (Facebook and LinkedIn), social news (Digg and Reddit), media sharing (YouTube and Instagram), blogs (WordPress and Blogger), microblogs (Tumblr and Twitter), review sites (TripAdvisor and Yelp), wikis (Wikihow and Wikipedia), social bookmarking (StumbleUpon and Delicious) and question-and-answer sites (Ask.com and Yahoo! Answers). (Gandomi and Haider, 2015, p. 142)

All of the previous mentioned techniques also apply in social media analytics. However, a few techniques only apply to the social media. These are community detection, link prediction and social influence measurements.

#### 3.1.5.6.1 - Community detection

Community detection is a big data technique, which extract and locate communities within a larger online social network. A community refers to a group of users that interact more extensively with each other than with the rest of the network. (Gandomi and Haider, 2015, p. 142)

In marketing, analysts use community detection to develop a more efficient product recommendation system. The World Wide Web use community detection to locate web clients who are geographically near and have similar interests. Thereby, a dedicated proxy can more efficiently serve the community. Organizations use community detection to locate and analyze communities to receive consumer feedback indirectly. This is a great opportunity to analyze user-generated content. (Parthasarathy et al., 2011, p. 83)

According to Parthasarathy et al. (2011), "*At the most fundamental level, community discovery (either in a static or evolutionary context) can facilitate and aid in our understanding of a social system*" (Parthasarathy et al., 2011, p. 83). There can be millions of nodes within a network and by the help of community detection techniques; it is possible to break the network into smaller communities. This happens by identifying clusters of people that have similar connections and interact with each other. Subsequently, the techniques can help outline behavioral patterns and predict emergent properties about the community. This enables researchers to get a richer understanding of the underlying social phenomenon. (Parthasarathy et al., 2011, p. 83)

#### 3.1.5.6.2 - Link prediction

Business intelligence apply link prediction techniques to reveal potentially connected criminal networks or terrorists. In biology, to discover associations in biological networks such as protein interactions. Online, analysts use link prediction as a recommendation system to suggest friends, movies or products. The technique analyzes the already established network in order to predict additional links. Within the social media, it would typically involve analyzing connections between shared friends or shared interests. As the social networks are not static but continuously shifting, new links will naturally appear over time. Therefore, it is a natural aim to understand and predict the more underlying dynamic of networks such as the occurrence of collaboration, interaction and influence between nodes. (Gandomi and Haider, 2015, p. 143)

#### 3.1.5.6.3 - Social influence detection

The behavior of an individual affects by the people in his network, and naturally, some people have more influence than others do. Social influence techniques attempt to quantify and model the social influence of actors and the strength of connections in a social network. These techniques are common within marketing, and can measure the marketing value of bloggers and famous people. The two most well-known models for evaluating the influence of an actor are the Linear Threshold Model (LTM) and the Independent Cascade Model (ICM). (Gandomi and Haider, 2015, p. 142)

The ICM focus on individual influence and interaction, while the LTM put an emphasis on collective influence. These models describe the theoretical foundation, but in practice, they are far simpler. (Chen et al., 2010, p. 88)

The two models are based on two different aspects on how social interaction works. For both models, each node represents an individual and the edge represent the relationship between them. The node is either active or closed, and if it is active, it means that it has been influence by the relationship it shares with the other nodes. In the Linear Threshold Model each edge has a weight, which refers to the influencing power one person have over another. Furthermore, each node has a required value that makes it activate if the value exceeds the required value. When the sum of a node's edges is equal or above its value, then it becomes active. The social influence of a single node is the sum of all the nodes it can activate in the social network. (Chen et al., 2010, p.88)

In the independent Cascade model, there is no value that determine whether a node activates. Instead, a statistical approach determines if a node activates. The edge has an activation probability, which determines if the neighboring nodes activates. Every time a node activates, there is a probability that it will activate the neighboring nodes based on their relationship. Once more, the social influence of a node is the sum of all the nodes it has a chance to activate in a social network. It is necessary to calculate for the statistical probabilities in this model. (Chen et al., 2010, p.88)

## 3.2 - The Crisis Communication Field

### 3.2.1 - Coombs as an entry to the crisis communication field

The purpose of this part is to bring insights into how this paper apply W. Timothy Coombs' theoretical crisis communication & management framework, and why this framework was chosen to provide a theoretical entry to this field. The first section explains exactly why this paper apply Coombs, while the second section go into details with the important assumptions in the framework. The second section also introduces Coombs' definition of organizational crisis. The third section elaborates on the overall structures of the framework, which entails brief explanations of important models, tools, concepts and so on. The fourth section explains some important delimitations made in the preliminary analysis of the framework, while the fifth, and last, section goes deeper into the five areas that were found through the preliminary analysis. These five areas are the areas that part two of the analysis take under investigation. The fifth section explains what these areas are about, why they are relevant to examine, and where they fit into Coombs' framework.

### 3.2.2 - Why Coombs?

There are different reasons to why this paper applies W. Timothy Coombs' framework from the book 'Ongoing Crisis Communication: Planning, managing & responding' as the theoretical perspective on the crisis communication field. One contributing factor is that Coombs has become a dominant scholar within the field of crisis communication. According to Google Scholar, Coombs is by a distance the most cited academic within this field ("Google Scholar," n.d.) . This feeds the argument that Coombs has great academic influence within crisis communication and management. Out of several publications, the book 'Ongoing Crisis Communication: Planning, managing & responding' is by far his most cited piece of work ("Timothy Coombs - Google Scholar," n.d.). This makes the book initially interesting.

Another specific reason to why this book is particularly interesting is that the book offers a comprehensive framework about crisis communication and management. This framework provides a structural basis for a theoretical analysis through its ability to map out concepts, areas and topics within the field. This instantiates the approach to the field. Furthermore, the framework is a modern and representative depiction of the field in its current state. The arguments for this is that 1) the applied edition of the book is from 2015, 2) the book is up-to-date with reflections about the newest trends in the field concerning social media, impact of the internet etc. and 3) that the book utilizes academic research from many different scholars of the crisis communication field (Coombs, 2015, p. xi).

An additional argument for why this framework is a good choice is that it caters to both academics and practitioners (Coombs, 2015, p.xi+1). As this paper attempts to find the potential qualities of big data analytics in the crisis communication field this is beneficial. It is beneficial because it improves the conditions of treating whether it is possible to use big data analytics in praxis in relation to crisis communication and management. Since this paper operates on a theoretical level, crisis theory that contemplates practical use provide the best possible foundation for the analysis. This sort of theoretical material forms the best possible picture on how crisis communication and management works in praxis, and it is therefore easier to compare how practitioners of this field could integrate a practical discipline such as big data analytics into their current methods and work patterns.

### 3.2.3 - Important assumptions and a definition of organizational crisis

The choice of utilizing Coombs' framework in this paper entails a natural acknowledgement or acceptance of certain theoretical assumptions about what crisis communication and management is. This part clarifies some of the most important assumptions in the Coombs' framework. As some of these assumptions closely connects to Coombs' definition of organizational crisis, this part also introduces Coombs' definition of organizational crisis along with assumptions related to the definition.

Coombs buys into the idea that all crises have a life cycle (Coombs, 2015, p.6). Coombs expresses that the life cycle idea is common within crisis management literature (Coombs, 2015, p.6). The assumption that all crises have a life cycle is important because it constitutes the foundation for Coombs' framework. Coombs builds his framework on a specific life cycle model referred to as the three-stage approach (Coombs, 2015, p.9). The three-stage approach is, like other life cycle models, a conceptualization of how the life cycle of a crisis appears. The three stages in this approach are *pre-crisis*, *crisis* and *post-crisis* (Coombs, 2015, p.10-11). This model builds on the assumption that every organization either stands before a crisis, are undergoing a crisis or just came out of one. This assumption is important because it entails that organizations must always expect a crisis to break out at some point (Coombs, 2015, p.3). For this reason, the probability of crises occurring can never be totally eliminated (Coombs, 2015, p.3).

This leads to another important assumption that the crisis management discipline is an ongoing process (Coombs, 2015, p.1). Coombs makes it clear that the word process is a keyword in crisis management (Coombs, 2015, p.x). According to Coombs, process is a keyword because it clarifies that crisis management is a proactive discipline and not a reactive one (Coombs, 2015, p.x). Due to the life cycle dynamics of crises, the environment around an organization is under constant transformation. Having a crisis management plan is insufficient if the organization does not understand the need to change the plan accordingly to the environmental dynamics (Coombs, 2015, p.x). This count for all actions related to crisis management. Hence, organizations need to become aware about the process perspective and work proactively in order to manage organizational crises in the best possible manner.

### 3.2.3.1 - Coombs definition of organizational crisis

Certain assumptions are inseparable from Coombs' definition of organizational crisis, and it therefore makes sense to present these assumptions along with the definition. This section treats these assumptions along with Coombs' definition of organizational crisis, and explain how this crisis definition contribute to the paper.

Coombs defines organizational crisis as *"… the perception of an unpredictable event that threatens important expectancies of stakeholders related to health, safety, environmental, and economic issues, and can seriously impact an organization's performance, and generate negative outcomes"* (Coombs, 2015, p.3). In his elaboration of this definition, Coombs focuses mainly on two assumptions. The first is that organizational crises are perceptual, and the second is that organizational crises only develops through perceptions of stakeholders (Coombs, 2015, p.3). The second assumption somehow builds upon the first assumption, and both assumptions aligns with his definition of organizational crisis. What is important about these assumptions is the assertion that it is the perceptions of stakeholders in the organizational environment that decides whether an event is an actual crisis (Coombs, 2015, p.3). A crisis only exists if stakeholders believe that the crisis exists. Hence, organizational crises occur in the relationship between the organization and its stakeholders (Coombs, 2015, p.4).

The definition also implies some characteristics about organizational crises. According to Coombs, organizational crises are unpredictable and must have a potential to create negative outcomes or damage to the organization (Coombs, 2015, p.3-4). These characteristics are important because the description of organizational crises becomes more distinct. Events that are either too predictable or not in any way disruptive to the organization cannot classify as an organizational crisis.

The contribution of the organizational crisis definition is just as well to create a clear distinction about what an organizational crisis is not. Coombs acknowledges that other types of crises exists, and he differs between disasters and organizational crisis (Coombs, 2015, p.2-3). An organizational crisis is about the organization. Disasters are events that are disruptive on a much larger scale and they involves multiple governmental units (Coombs, 2015, p.3). A disaster can turn into an organizational crisis, but as long as the disaster does not concern or affect the organization, it is not an organizational crisis (Coombs, p.3). Coombs' definition of the organizational crisis helps to emphasize the fact that this paper only focuses on the organizational level in relation to crisis communication and management. In addition, the definition contributes with a fundamental perspective on what an organizational crisis is, which helps this paper approach the crisis phenomenon in an organizational context.

## 3.2.4 - Further elaboration on Coombs framework

The previous explanation of the three-stage approach slightly touched upon the overall structures of Coombs' framework. However, the approach merely constitutes the foundation on which Coombs builds his framework. The framework consists of sub-stages, models, concepts etc. that all together forms an extensive framework for crisis communication and management. A thorough understanding of the framework therefore requires further elaboration. However, an into depth explanation of the entire framework is both too extensive and unnecessary to the purpose of this paper. Still, it seems necessary to elaborate on the general structures of the framework. The main reason is that the last parts of this theoretical section focus on the specific areas in Coombs' framework that are in particular relevant to this paper. An over-all elaboration of the framework contributes with a contextual understanding about how these selected areas fit into the framework.

As mentioned in the previous section, the three-stage approach is a conceptualization of the crisis life cycle perspective. Coombs use this conceptualization as a sort of foundation or organizing structure to build his framework upon. This allow Coombs to group and introduce different sub-stages, tools, concepts, recommendations etc. in relation to pre-crisis, crisis and post-crisis concerns. The sub-stages within each of these crisis stages are the most important to understand because they outline the different areas connected to each crisis stage. The following sections elaborate on the sub-stages and general content within each of the three crisis-stages.

### 3.2.4.1 - The pre-crisis stage

The three subs-stages in the pre-crisis stage are signal detection, correction and crisis preparation (Coombs, 2015, p.10+44). The first two combined also defines as the crisis prevention process (Coombs, 2015, p.44). The crisis prevention process is an proactive management process that aim to avoid crises (Coombs, 2015, p.44). Signal detection is defined as the search for crisis warning signs, whilst correction is about the reduction or elimination of threats (Coombs, 2015, p.31+44). Coombs introduce the crisis prevention process through his five-step prevention model (Coombs, 2015, p.44). The steps in this model are 1) identifying the sources to scan, 2) collecting the information, 3) analyzing the collected information, 4) preventive actions if necessary, and 5) evaluating the effectiveness of the threat reduction initiatives (Coombs, 2015, p.44). The first three steps define the signal detection phase, whilst the last two steps define the correction phase (Coombs, 2015, p.44). The first three steps in this five-step model is also referred to as the crisis-sensing mechanism (CSM) (Coombs, 2015, p.58).

Coombs suggests that crisis practitioners look towards established proactive management functions when designing a CSM, in order to save money, time and effort (Coombs, 2015, p.31). The suggested management functions are risk management, issues management and reputation management (Coombs, 2015, p.31). The purpose of the signal detection phase is to scan and monitor for potential internal and external crises, so that the organization is able to activate the correction phase when necessary (Coombs, 2015, p.31). Coombs mentions these three proactive management functions because they possess tools, sources and methods to do this (Coombs, 2015, p.31).

The crisis preparation process has a different focus than the crisis prevention process. Whereas the crisis prevention process focuses more on detecting and acting on crisis warning signs, the crisis preparation process focus on the necessary preparations organizations should conduct. The crisis preparation process builds on the assumption that organizations cannot prevent every crisis from occurring. A crisis will break out at one point, and organizations must therefore prepare for crisis-breakouts as well (Coombs, 2015, p.66). Similar to the five-step prevention model, Coombs introduces a six-step model for crisis preparation (Coombs, 2015, p.66). The steps in this model are 1) diagnosing vulnerabilities, 2) assessing crisis types, 3) selecting and training a crisis management team, 4) selecting and training spokespersons, 5) developing a crisis management plan (CMP), and 6) reviewing the crisis communication system (Coombs, 2015, p.66). This model includes all the different preparations and installments Coombs believes are required to make an organization prepared.

#### 3.2.4.2 - The crisis stage

Coombs believes that any crisis event separates into two sub-stages or crisis phases. He defines the first phase as the crisis recognition phase and the second phase as the crisis response phase (or the crisis containment phase) (Coombs, 2015, p.11). In the crisis recognition phase, Coombs describes how information processing is a central part of crisis management when a crisis breaks out (Coombs, 2015, p.118). Coombs perceive the beginning of a crisis event as an information-poor and knowledge-poor situation (Coombs, 2015, p.118). Many things are unknown to the organization (Coombs, 2015, p.118). According to Coombs, crisis practitioners should transform the situation from ‘unknown’ to ‘known’, so that the organization can work effectively in terms of solving the crisis and figure out the extent of the crisis (Coombs, 2015, p.118). Coombs uses the term ‘situation awareness’ to describe this process (Coombs, 2015, p.118).

In relation to the information-processing phase, Coombs highlights some current tools that can aid crisis practitioners to achieve an effective process. Some structural elements he suggests are crisis information logs and a crisis knowledge maps (Coombs, 2015, p.123-125). A crisis information log is a log that track the amount and movement crisis-related knowledge and information inside the organization (Coombs, 2015, p.124). A crisis knowledge map is a map over the network that pinpoints which internal and external stakeholders that can provide necessary information in the time of crisis (Coombs, 2015, p.123). On top of this, Coombs also suggest some procedural elements such as categorization of information and data splitting (Coombs, 2015, p.125-126). Categorizing information helps the practitioners focus on top priority information, while data splitting help combat information acquisition biases through dividing the data randomly between practitioners working on the crisis event (Coombs, 2015, p.125-126). The purpose of the structural elements is to develop a structure around how to collect information and manage the current knowledge in the organization, while the purpose of the procedural elements is to prevent or reduce information-processing errors (Coombs, 2015, p.123).

When a crisis is recognized, the organization must activate the crisis response phase, which entails performance of effective crisis communication that a) prevent the crisis from spreading, and B) limit its duration (Coombs, 2015, p. 129-130). According to Coombs, there are three areas within crisis communication in connection to the crisis response. These are form, strategy and content (Coombs, 2015, p.130). Form focus on the tactical aspect of the crisis response, and Coombs main advice is that organizations must respond fast, speak in one voice and be as open as possible about the crisis situation (Coombs, 2015, p.129-134). Strategy is about having a strategic focus on the crisis communication, and determine what the crisis communication hope to accomplish (Coombs, 2015, p.136). In this context, Coombs suggest that the organization develop objectives and determining the target audience of the crisis response (Coombs, 2015, p.136-138). Content includes 1) instructing information, 2) adjusting information, and 3) reputation management and repairing (Coombs, 2015, p.139). These three areas refer to the actual content of the crisis response.

In terms of actual crisis response tools, the most significant in Coombs framework is the crisis response strategies and the SCCT framework that he presents. SCCT stands for *situational crisis* *communication theory*, and the framework helps organize the crisis response strategies and determine which response strategy to choose based on the intent of the response (Coombs, 2015, p.144+146). Coombs introduces 10 different crisis response strategies (Coombs, 2015, p.147). The choice of strategy depends on the reputational threat of the crisis (Coombs, 2015, p.150). Coombs suggests the use of the SCCT to evaluate the reputational threat (Coombs, 2015, p.150). The three evaluation factors in the SCCT are 1) the crisis type, 2) crisis history, and 3) prior reputation (Coombs, 2015, p.150). Determining the crisis type is the first step in evaluating the reputational threat. The crisis type is decisive for the level of responsibility (Coombs, 2015, p.150). Responsibility for the crisis is important to know, because high responsibility is a larger threat to an organization's reputation than low responsibility (Coombs, 2015, p.151). Coombs provides a palette of 10 different crisis types split into three different levels of responsibility (Coombs, 2015, p.150). After having assessed the level of responsibility, the organization must consider their crisis history and reputation (Coombs, 2015, p.151). If the organization have had similar crises before it is likely to cause more damage on the reputation and create a *Velcro effect* (Coombs, 2015, p.151). Moreover, if the organization currently have a bad reputation it is more likely that stakeholders will put more responsibility on the organization then they would otherwise (Coombs, 2015, p.151).

### 3.2.4.3 - The post-crisis stage

The post-crisis stage does not consist of any sub-stages, but the framework still contains three main areas of post-crisis concerns that explains the important tasks and responsibilities of practitioners after a crisis. These three areas are crisis evaluation, memory & learning, and post-crisis actions (Coombs, 2015, p.162). Crisis evaluation entails the performance of two different evaluations (Coombs, 2015, p.162). The first evaluation concerns the crisis management performance. This evaluation goes into details with the performance of the organization and crisis management team (CMT) on a number of crisis related areas (Coombs, 2015, p.164). These areas include how well the CMT did throughout the different crisis phases, how well the different systems (technical systems, infrastructure, CMP etc.) functioned throughout the crisis, how the stakeholders have reacted to the performance of the organization, and how well the organization did with this specific crisis type (Coombs, 2015, p.164-165).

The second evaluation is an impact evaluation (Coombs, 2015, p.166). This evaluation is about assessing the damage caused by the crisis event. Coombs suggest that the organization could use the strategic objectives that the organization developed in the crisis response phase as specific measurements (Coombs, 2015, p.136+166). These objectives include both stakeholder-related objectives, such as the psychical safety and psychological well-being of stakeholders, and organizational-related objectives (Coombs, 2015, p.166). The most central organizational-related objectives include reputation and financial performance (Coombs, 2015, p.166). Coombs believe that a comparison of pre- and post-crisis reputations is the strongest indicator of the reputational impact of the crisis (Coombs, 2015, p.166). Coombs highlights the importance of social media, the internet, stakeholder feedback and traditional media as sources to determine the reputational impact (Coombs, 2015, p.166-168). Financial performance is another source to estimate impact. Sales, market share, stock price etc. provides another perspective on crisis impact, and a comparison of the pre- and post-crisis financial figures provides a strong indication on that damage caused by the crisis (Coombs, 2015, p.167).

Both memory & learning and post-crisis actions are only briefly discussed by Coombs (Coombs, 2015, p.170-175). The key takeaways from the memory & learning part are that organizations must store crisis documentation and evaluation reports for later use and make the documents retrievable, as well as utilizing these documents to learn about past mistakes (Coombs, 2015, p.170-171). In terms of post-crisis actions, Coombs mentions three tasks of the CMT in the post-crisis stage. These tasks are cooperation with investigations, follow-up communication and crisis tracking (Coombs, 2015, p.172). Cooperation with any necessary investigations aim to build goodwill with authorities, and indicates an open and honest approach to the crisis in front of the stakeholders (Coombs, 2015, p.172). Follow-up communication rebuilds relationships with the stakeholders and try to prevent the crisis from returning (Coombs, 2015, p.172-173). Crisis tracking monitor whether there are any threats of the crisis breaking out again (Coombs, 2015, p.175).

### 3.2.5 - Elaboration of the areas found in the preliminary analysis

The methodological part of this paper explained the preliminary work leading up to the analysis. Just to summarize, a part of this preliminary work entailed an analysis of Coombs' framework with the purpose of finding areas within the framework that includes data collection and analysis. The reason behind this preliminary analysis is to find areas within the framework that might have an increased potential in terms of applying methods from big data analytics. The preliminary analysis found five areas of interest. These five areas are expectation gaps, stakeholder salience, situation awareness, social media monitoring during a crisis, and crisis evaluation. The purpose of the following section is to provide a better understanding about each of these areas. This section focuses on what each of the areas are about, why they are relevant, and how they fit into the overall framework.

### 3.2.5.1 - Expectation gaps

Coombs presents the concept of expectation gaps as a part of the crisis prevention process (Coombs, 2015, p.55). An expectation gap is any gap between stakeholders' expectations and the organizational actions (Coombs, 2015, p.56). When crisis practitioners analyze the likelihood and impact of potential reputational crises, Coombs points out that it is first necessary to determine whether any expectation gaps exist (Coombs, 2015, p.55). The reason why this is important is that any expectation gap pose a reputation threat, and can lead to a reputational crisis (Coombs, 2015, p.55). The first step is to find stakeholders expectations (Coombs, 2015, p.55). Different stakeholders have different expectations, and Coombs suggest that organizations therefore mainly focus on the major stakeholder groups (Coombs, 2015, p.55). When the expectations are found the next step is to determine whether stakeholder perceive that the organization live up to these expectations (Coombs, 2015, p.55). Coombs explains that there are two types of expectation gaps, and these are called performance gaps and perceptions gaps (Coombs, 2015, p.55-56). Performance gaps exists when the organizational actions do not match the stakeholder expectations (Coombs, 2015, p.56). Perception gaps exists when organizational actions actually match the stakeholder expectations, but the stakeholders are unaware of it (Coombs, 2015, p.56). Perception is a very important concept in connection to expectation gaps. As the perception gap definition shows, it is the perception that determines whether a gap exists (Coombs, 2015, p.55). It does not matter whether the organization actually match stakeholder expectations or not, if stakeholders perceive it differently.

Coombs explains that stakeholders build their perception about the organization on direct and indirect interactions (Coombs, 2015, p.35). Direct interactions take place when the stakeholders have experiences with the organization or their services (Coombs, 2015, p.35). Indirect interactions take places through news reports, social media, friends, family etc. (Coombs, 2015, p.35). The interaction result in a positive or negative perception of the organization (Coombs, 2015, p.35). As the stakeholder experience multiple interactions with the organization, each interaction contribute to this perception in either a positive or negative manner, depending on whether each interaction exceeds or disappoints compared to the stakeholder's expectation (Coombs, 2015, p.35). When a stakeholder expresses how he perceives an organization, the sentiment of the communication depends on the stakeholder's individual expectation gap. Based on the suggestion that there exist both favorable and unfavorable reputations, and that *"reputations are formed as stakeholders evaluate organizations based in direct and indirect interactions"* (Coombs, 2015, p.35), expectation gaps can take form as both positive and negative (Coombs, 2015, p.34-35).

The main reason why it is relevant to look at expectation gaps is that the assessment of these is instrumental for detecting and preventing reputational crises. If big data analytics have any potential to detect expectation gaps, it might benefit crisis practitioners in the crisis prevention phase.

### 3.2.5.2 - Stakeholder salience

The stakeholder salience concept belongs to the same area as expectation gaps. When a crisis practitioner finds an expectation gap, it is important for him to evaluate the stakeholder's salience to determine the likelihood and impact of the potential reputational crisis that the expectation gap poses (Coombs, 2015, p.55). The stakeholder salience concept expresses the actual importance of the stakeholder to organization (Coombs, 2015, p.55). The stakeholder salience concept is a function of three components. These are power, legitimacy and willingness, and it is possible to convert them into likelihood and impact scores (Coombs, 215, p.55). The power component covers a stakeholder's ability to make an organization do something it would not do otherwise (Coombs, 2015, p.56). Power develops through a couple of different factors such as resources (e.g. money, contacts and communication channels) and coalition formations (Coombs, 2015, p.56-57). If stakeholders create a coalition they are often more powerful than single-handedly.

Legitimacy refers to stakeholder actions that other stakeholders may consider a legitimate action due to its appropriateness or desirability (Coombs, 2015, p.57). If other stakeholders consider a concern by a stakeholder to be valid, this concern will gain legitimacy in the stakeholder environment (Coombs, 2015, p.57). Coombs warns that if the organization ignores this it will *"... make the organization appear callous to the other stakeholders"* (Coombs, 2015, p.57). This increase the risk that the concern will spread and damage other stakeholder relationships (Coombs, 2015, p.57). Finally, willingness refers to the likelihood that a certain stakeholder will confront the organization (Coombs, 2015, p.57). If a problem is important to the stakeholder or the relationship between the organization and the stakeholder is weak, it is more likely that the stakeholder will confront the organization (Coombs, 2015, p.57).

Coombs argues that it is possible to convert power, legitimacy and willingness into likelihood and impact scores for a reputational threat becoming a reputational crisis (Coombs, 2015, p.57). Power and legitimacy combined constitute the impact score, while legitimacy and willingness constitute the likelihood score (Coombs, 2015, p.58). The reason why stakeholder salience is relevant to look at is that Coombs does not provide any specific methods to collect data about power, legitimacy and willingness. Rather it seems that a subjective evaluation determines the scores. The crisis practitioner, however, still needs to collect and base the analysis on valid information about stakeholders, and it is therefore interesting to explore whether big data analytics is able to contribute within anything within this area.

### 3.2.5.3 - Situation awareness

Both the expectation gaps and the stakeholder salience are parts of the crisis prevention process. Situation awareness is on the other hand a part of the crisis recognition phase (Coombs, 2015, p.118). The elaboration of Coombs' framework already touched slightly upon this concept, and it described the concept as a term to describe the process of transforming a situation from "unknown" to "known". A more detailed explanation of situation awareness is that it involves three different areas. These are the perceptions of the situation and environment, the comprehension of them, and the ability to project future states (Coombs, 2015, p.118). Situation awareness describe a state in which the crisis practitioners obtain a perceptual understanding of a crisis situation, and based on this can predict the effects of the crisis and determine what necessary actions the organization need to take in order to cope with the crisis (Coombs, 2015, p.118).

As the prior elaboration of the framework mentioned, the beginning of a crisis event is a knowledge-poor and information-poor situation. When only little is known, such a situation requires a collection of large amounts of information to determine the existence and the extent of the crisis (Coombs, 2015, p.118-119). When a situation is unknown, Coombs recommends that the first three tasks of crisis practitioners is to figure out what they need to know, what they already know and what they do not know (Coombs, 2015, p.119). These the first steps towards the creation of situation awareness. After practitioners have made the initials steps toward situation awareness, they must begin the information-gathering phase (Coombs, 2015, p.119).

Coombs links situation awareness to information gathering and information processing in the crisis recognition phase (Coombs, 2015, p.118-119). Coombs describes information gathering as an organized search for information (Coombs, 2015, p.119). This entails that the organization must prioritize their information needs and develop strategies about how to get this information (Coombs, 2015, p.119). Coombs deems the aforementioned crisis knowledge maps as useful tools to assist practitioners in the information gathering process, because a crisis knowledge map helps practitioners to figure out where to get different information (Coombs, 2015, p.119). The information-processing concept covers the actual process of making sense out of the collected information (Coombs, 2015, p.120). The main purpose of the process is for practitioners to analyze information and determine whether they have a sufficient amount of information to make effective decision (Coombs, 2015, p.120). The previous elaboration of the framework already clarifies some structural and procedural elements within information processing.

A few things make it relevant to investigate situation awareness further. Primarily it is relevant due to the elements of data collection and analysis. The transformation from an "unknown" to a "known" situation entails extensive collection and analysis of data, and it is interesting to see if big data analytics can contribute to this process. In addition, as Coombs explains the importance of stakeholder perception and the projection of future states, it is interesting to see whether big data analytics can contribute to with any collection or analysis methods within these areas. Coombs emphasizes stakeholder relationships as the main data source in the situation awareness process.

### 3.2.5.4 - Social media monitoring during a crisis

The monitoring of social media during a crisis is not an established area in Coombs' framework compared to the other areas that the preliminary analysis found. This area is rather a part of Coombs' considerations about social media when an organization is in a crisis (Coombs, 2015, p.155-157). This might be because it is a new area in crisis communication and management. Still, Coombs emphasizes that social media monitoring remains essential during a crisis (Coombs, 2015, p.157). Social media monitoring during a crisis entails that crisis practitioners must look out for what stakeholders are saying, whether the comments are favorable or unfavorable and what the dominant topics are among stakeholders during a crisis (Coombs, 2015, p.157). This might be a difficult task since the large amount of different social media increase the amount of online communication channels (Coombs, 2015, p.18+21). In addition, social media also increase the number of potential "crisis voices", which only makes the monitoring more complex (Coombs, 2015, p.156).

The fact that Coombs considers social media monitoring essential during a crisis is the main argument for why this paper look into this area. In addition, this paper interprets social media monitoring during a crisis as a search for important crisis-related information. When a crisis practitioner needs information there is a need for collection and analysis of data. This paper therefore sees a potential connection between the abilities of big data analytics and social media monitoring during a crisis.

### 3.2.5.5 – Crisis impact evaluation

The last area found in the preliminary analysis is crisis evaluation. The previous elaboration of Coombs' framework clarified that two types of crisis evaluations are important in the post-crisis stage. Just as a reminder, the two evaluations are the crisis management performance evaluation and the impact evaluation. The first evaluation goes into detail with the performance of the organization and crisis management team, and the second evaluation goes into details with the damage or impact caused by the crisis (Coombs, 2015, p.164+166).

The crisis evaluation area is interesting because data collection and analysis is fundamental in both of these evaluation processes (Coombs, 2015, p.163+167). In terms the crisis management performance evaluation, Coombs suggests that practitioners should collect data from sources such as crisis records, stakeholder feedback, organizational performance measures, internet comments, and media coverage (Coombs, 2015, p.163). The crisis records are the most useful data source for evaluating how the crisis management team (CMT) worked and performed internally during the crisis (Coombs, 2015, p.163). The stakeholder feedback, internet comments and media coverage are handier for evaluating the perceived crisis management performance by stakeholders (Coombs, 2015, p.163). Both the internal and external evaluation of the crisis management performance are equally important (Coombs, p.163). In this context, this external evaluation is especially relevant due to the digital media data sources. The external evaluation of the crisis management performance seeks to determine whether specific crisis management actions by the organization had an effective or ineffective impact, and whether the stakeholders feel that the organization handled the crisis situation well (Coombs, 2015, p.165). According to Coombs, organizations usually gather this feedback through surveys, interviews and focus groups, but because cooperation can be difficult with certain stakeholders after a crisis, he suggests the use of the online media (Coombs, 2015, p.163). It is therefore interesting to see if big data analytics can contribute in this context.

A way to perceive the impact evaluation is that it is an extension of the crisis management performance evaluation specifically focused on the impact of the crisis (Coombs, 2015, p.166). As the previous elaboration clarified, the focus of this evaluation is reputational damage and financial damage. Coombs puts a higher emphasis on digital media as data sources compared to the data sources for the crisis management performance evaluation (Coombs, 2015, p.166-167). A central part of the reputation evaluation is look for positive and negative responses in traditional and digital media, and track these by time to indicate the gradient development these responses (Coombs, 2015, p.167). Also, it is important to track the “word-of-mouth” on social media to see if stakeholder is talking positively or negatively about the organization after a crisis, and what stakeholders like or do not like about the organization (Coombs, 2015, p.167). In terms of financial damage, Coombs suggests that practitioners can track financial data such as sales, markets share, stock prices etc. to evaluate the impact of the crisis on the organization’s financial performance (Coombs, 2015, p.167).

## 3.3 - Luhmann: Systems theory

### 3.3.1 - Introducing Luhmann in this paper

This paper use fragments from Luhmann's systems theory to gain perspective on communication and complexity. The analysis mainly applies these fragments when analyzing the capabilities of big data analytics. Luhmann provides the possibility to apply a different perspective on big data analytics, and he allow the analysis to take a step back, and consider the actual contributions of this technology in relation to social life. Luhmann's theoretical work is complex, and it is needless to explain all the concepts in his systems theory. Instead, the following sections make a brief introduction to the relevant areas in his theory that this paper draws upon.

### 3.3.2 - The systemic perspective

The ultimate assumption in systems theory is that systems exist (Luhmann, 1995, p.12). The paradigm shift from studying individual phenomena to studying systems (the network and relations between phenomena), derives from a clash in scientific approaches between biology and psychics (Kneer and Nassehi, 1997, p.26). Biologists dispute that psychical and chemical processes of organisms alone can explain life, as organisms never appear isolated, but in relation with other organisms (Kneer and Nassehi, 1997, p.26). Hence, 'the system concept' refers to networks of organisms and a scientific approach of examining the interrelationships between these organisms, rather than exploring the compositions of the single units in a system.

Niklas Luhmann marks the adoption of systems theory in social science through his theory of social systems. Biological systems (organisms) are only one of four branches in general systems theory from his perspective (Luhmann, 1995, p.2). The other three are social systems, psychic systems and mechanical systems (Luhmann, 1995, p.2; Thyssen, 2012, p.690). Social systems are the central area in Luhmann's work, though he also treats the other types of systems. Through the development of social systems theory, Luhmann attempts to formulate a universal theory for sociology that combines the field with the methods of general systems theory (Kneer and Nassehi, 1997, p.37; Luhmann, 1995, p.15). Luhmann defines general systems theory as a 'super-theory' due to its claim of universality (Luhmann, 1995, p.4). The attempt to combine sociology with other sciences through a super-theory, generates a meta perspective to explain social life in context with other areas (Thyssen, 2012, p.686).

### 3.3.3 - Systems as autopoietic systems

Luhmann differs between the mechanical systems and the autopoietic systems (Thyssen, 2012, p.690). Autopoiesis refers to the biological process of self-reproduction, and autopoietic systems are hence systems that reproduce themselves (Hornstrup, 2005, p.15; Kneer and Nassehi, 1997, p.52). Both the biological systems, psychic systems and social systems belong in this category (Thyssen, 2012, p.690). Apart from being self-reproducing, autopoietic systems are operationally closed, meaning that they neither have a input or output like a mechanical system have (Kneer and Nassehi, 1997, p.60; Thyssen, 2012, p.694). Mechanical systems are reliable in the sense that they produce a predictable output based on the input (Thyssen, 2012, p.694). Autopoietic systems are neither predictable or producers of an output (Kneer and Nassehi, 1997, p.60; Thyssen, 2012, p.694). Instead, they are self-referential through their ability to establish relations with themselves, and differentiate those relations from relations with the environment (Luhmann, 1995, p.13).

Autopoietic systems reproduce themselves within the boundaries of their medium (Thyssen, 2012, p.694). The boundary itself is what defines the system, as it marks the difference between the system and its environment (Luhmann, 1995, p.28-29; Thyssen, 2012, p.695). But also, the boundaries are what connects the system with its environment (Luhmann, 1995, p.28). The boundary separates elements and events of systems, but allow causal relations (Luhmann, 1995, p.29).

### 3.3.4 - Social & psychic systems – and communication

The medium of an autopoietic system is imperative, because it is "the body" of the system (Thyssen, 2012, p.694). Social systems operate within the medium of communication, whilst psychic systems operate within the medium of consciousness (Thyssen, 2012, p.694). Both consciousness and communication operates through different devices, but are not equal to their devices (Kneer and Nassehi, 1997, p.66). The devices only permit the mediums to exist, but is not the medium themselves. Consciousness operates through the brain and brain processes, while communication operates through different communication devices that makes it easier to communicate (Thyssen, 2012, p.691; Luhmann, 1995, p.162; Kneer and Nassehi, 1997, p.66). The most fundamental communication device is language (Thyssen, 2012, p.691). The purpose of language is to solve any problems concerning the 'understanding' of the communication (Thyssen, 2012, p.691). Communication is always an uncertain action (Thyssen, 2012, p.691). This is because it is impossible to know whether the interpretation of the communication aligns with connotation behind it (Thyssen, 2012, p.691). Language does not obliterate this problem, but it makes communication more sensible and understandable, which reduces the amount of misinterpretations (Thyssen, 2012, p.691).

Another type of communication device is a dissemination media (Thyssen, 2012, p.691). There are many different types of dissemination media. The concept covers writings, radio, television, internet etc. (Thyssen, 2012, p.691). The commonality of the dissemination media is that they preserve the communication (Thyssen, 2012, p.691). Communication is an occurrence, and it disappears instantly unless a dissemination preserves it (Thyssen, 2012, p.692). The dissemination media liberates communication from the demand of psychical presence when it takes place, and they make it possible to spread communication extensively (Thyssen, 2012, p.691). At last there is the functional systems contributes with dissemination media such as rhetoric and symbolic generalized media (Thyssen, 2012, p.691). These dissemination media generalize and simplify, which increase the effect of the communication (Thyssen, 2012, p.691). The functional systems are a chapter in itself, and this paper does not apply these systems in an analytical context. Therefore, this theoretical elaboration does not include further explanation of these.

In order for communication to take place, there must be at least two psychic systems present (Thyssen, 2012, p.690). Psychic systems and socials systems have evolved together through the development of meaning (Luhmann, 1995, p.59). Social systems cannot evolve without consciousness, and psychic systems cannot evolve without communication (Luhmann, 1995, p.59). Psychic systems takes part in the environment of social systems, like social systems take part in the environment of psychic systems (Luhmann, 1995, p.255). The systems are still separated and work within their respective mediums (Thyssen, 2012, p.696). However, the systems are able to present themselves sensible to each other and offer their individual complexity in a simplified form (Thyssen, 2012, p.697).

A social system is a zone of meaningful context based on social interactions that refers to each other (Kneer and Nassehi, 1997, p.50). A psychic system is the same, but it consists of thoughts or conceptions instead of social interactions (Kneer and Nassehi, 1997, p.64). As due to the assertion that social systems only consists of communication, and that communication in itself reduces complexity, the interior of the social system is a reduction of complexity (Thyssen, 2012, p.690+694). Every type of social contact is understood as a social system (Luhmann, 1995, p.15). Hence, all social contact inherently reduces complexity as it works through social systems. Without the social systems, social interactions would not make any sense due to the absence of a meaningful context (Luhmann, 1995, p.62-63).

### 3.3.5 - Complexity

Complexity is a central concept in Luhmann's systems theory. Complexity defines as the total amount of possible events and shapes a system can take (Kneer and Nassehi, 1997, p.44; Luhmann, 1995, p.24). This 'infinity' is too complex for both the consciousness and the communication to grasp, and the systems cannot generate any meaning about the world in these conditions (Kneer and Nassehi, 1997, p.44). As a result, social and psychic systems attempt to reduce complexity through the exclusion of possibilities (Kneer and Nassehi, 1997, p.45). When excluding possibilities, it becomes more obvious what the "true meaning" is. However, the systems still combat complexity because they have to defend the exclusion of certain possibilities. In order to become resistant to the complexity of the world, systems have to develop their own complexity (Kneer and Nassehi, 1997, p.45; Luhmann, 1995, p.31). This generates two types of complexity. First, there is the environmental complexity, which is everything outside the boundaries of systems (Kneer and Nassehi, 1997, p.44). Second, there is the complexity of systems, which is the complexity within the boundaries of systems developed with the purpose of making the systems resistant to the environmental complexity (Kneer and Nassehi, 1997, p.45).

The complexity of systems have to sufficiently tackle and comprehend changes in global conditions (Kneer and Nassehi, 1997, p.46). If the system lack the complexity that allow it to perceive and process the complexity of the world, its existence is at stake (Kneer and Nassehi, 1997, p.46). The more complex a social system is, the better it can react to the challenges from its environment (Kneer and Nassehi, 1997, p.46). A social system can never become imperturbable, but it can become meta stabile (Thyssen, 2012, p.696). This means that it has developed the ability to resist change by incorporating change in its own structures.

## 3.4 - Entman: Framing theory

Erving Goffman first introduced the concept of framing in his book Frame Analysis (1974) and since then, various fields and practices have embraced the concept ("Framing Theory," 2011). Framing can help the paper to understand, which effect communication can have over the human consciousness. It brings light to how communication takes place and what the paper can derive by analyzing communication. A speech, news report or utterance can affect the way individuals perceive reality (Entman, 1993, p. 53). Therefore, it is a key concept in understanding how stakeholders perceive crises and how the phenomena arise. It helps the analysis understand how to apply big data techniques and comprehend the results.

### 3.4.1 - Definition of framing

Despite the omnipresence of the concept of framing across the humanities and social sciences, there is no clear description of how framing influence thinking (Entman, 1993, p. 51). Robert M. Entman (1993) argues there is a fractured paradigm of framing and it is necessary to synthesize the key concepts in order to construct a coherent theory (Entman, 1993, p. 51). As a result, he identifies the common tendencies among the various uses in order to create a more precise and universal understanding (Entman, 1993, p.51). Thus providing a composed framework of the key concepts in framing, which forms the basis for further research (Entman, 1993, p.51). This paper uses this framework as an understanding of the concept of framing.

According to Entman:

*"To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described."* (Entman, 1993, p. 52)

In this context, the interpretation of "*a communicating text*" is beyond written material and is valid for any transfer of information from one individual to another. Entman clarifies this, when he describes that framing involves any transfer of information, which include speech and utterance (Entman, 1993, p. 52). Frames highlight certain information about an item that is subject of communication and thereby making some information more meaningful and noticeable to the audience. (Entman, 1993, p. 53)

### 3.4.2 – Locations and functions of frames

According to Entman, there are four locations of frames in the communications process, which are the communicator, the text, the receiver and the culture. The communicators are not always aware of the framing process. Frequently, the communicators make unconscious framing judgements in deciding how to express themselves. Already established frames, guides the framing judgements and organize the audience's belief systems. (Entman, 1993, p.52)

The communicator manifests the frames in the text by the presence or absence of certain keywords, stereotyped images, sources of information, stock phrases or phrases that reinforces clusters of judgements and facts. The frames in the text guides the receiver and affect the conclusions drawn from the text. However, the reflections and the conclusions made by the receiver does not necessarily represent the frames in the text or the intended framing from the communicator. Culture is the last location, which describes the frequently applied frames in the discourse and the thinking of most individuals in a social grouping. The framing process in all locations may contain similar functions of selecting and highlighting certain elements to construct an argument about problems, their causation, evaluation and solution. (Entman, 1993, p.52)

To frame essentially contains one or more of four functions (Entman, 1993, p. 52):

1. Defining problems that involves determining the cost and benefits of an action from a causal agent. This is typically measured in common cultural values.
2. Diagnosing causes and identifying what forces are creating the problem.
3. Making moral judgements of the causal agent and the effects of the actions.
4. Suggesting remedies for the problem and predicting the likely effects.

Framing theory suggest that how the communication present something affects the way the audience perceive and process that information. It does not only influence what the audience think about but also how to think about this piece of information (Entman, 1993, p. 54). Kahneman and Tversky (1984) made an experiment to verify this point, which demonstrates the power of framing. The experiment involves numerous test subjects and the scientists asked them the following:

"*Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the program are as follows:*
*If Program A is adopted, 200 people will be saved. If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. Which of the two programs would you favour?*" (Tversky and Kahneman, 1981, p. 453)

Given the situation, 72% of the subjects selected Program A and only 28% selected Program B. In the subsequent experiment, the scientists gave the test subjects the same situation. However, the scientist framed the situation differently. They framed the information in terms of likely deaths instead of likely lives saved. If Program C was selected, 400 people would die. If Program D was adopted there is a 1/3 probability that 600 would die and 2/3 probability that no one would die. In this experiment, the results were completely opposite. Of all the test subjects, 22% choose Program C and 78% choose Program D. This experiment shows how the actions of a large portion of the audience changes through framing of information. (Tversky and Kahneman, 1981, p. 453)

Framing has the ability to change the way the audience perceive and act upon reality. The perception of a phenomenon can change radically depending on how the communicator frame a situation (Edelman, 1993, p. 232). As a result, the social reality is not to be understood as stable, but instead as a kaleidoscope of potential realities, which the communicator evokes through framing of information (Edelman, 1993, p. 232). Several different perceptions of reality can emanate depending on the framing of a potential crisis. Thus illustrating the influence of politicians, media and spokespersons in a crisis. Ultimately, framing plays a big part when a crisis unfolds and affects the way the stakeholders perceive the situation. A situation can easily become a huge crisis or be brushed off as something irrelevant (Coombs, 2015, p. 111). Therefore, it is essential for a crisis manager to understand how the stakeholders perceive the situation. Nevertheless, framing can also be a tool for the crisis practitioner to limit the consequences of a crisis (Coombs, 2015, p. 111).

# Chapter 4 - Analysis

This analysis consists of three different parts. Each part contributes differently to the process of answering the research question. The first part takes a step back and reflect on the actual capabilities of big data analytics in general. With inspiration from Luhmann and Entman, this part analyzes the general abilities of big data analytics. Concepts from Luhmann and Entman allow an analysis about the boundaries of big data analytics towards complexity, social dynamics and communication. This is important since it connects the general possibilities and limitations of big data analytics with complex social situations, such as organizational crises. When the analysis goes further into the process of answering the research question, reflections from the first part helps to explain the general capabilities of big data analytics. Knowing what big data analytics is capable of in general, helps determine if the technology can contribute with anything new to the existing methods and tools in the crisis communication & management area, and it aids the evaluation about the theoretical possibilities and limitations of big data analytics in crisis communication & management in particular.

The preliminary analysis of Coombs' framework found five particular areas interesting in relation to information gathering and analysis. The theoretical part about Coombs' framework elaborated on these five areas. The second part of this analysis takes these five areas under the microscope, and analyzes whether big data analytics can contribute with any technical improvements or new approaches compared to the current methods and tools of these areas. In this regard, the second part tries to figure out whether big data analytics can contribute to the existing methods in crisis communication & management. The approach to each of the five areas builds on a general three-step method. First, the analysis takes on the information-needs within each area. Second, the analysis then considers what exact big data techniques have a potential to assist a crisis practitioner when attempting to obtain this kind of information. Third, the analysis compares potential techniques with traditional methods and tools provided by Coombs, and analyze strengths and weaknesses of the big data techniques as well as the traditional methods. Through these comparisons, the paper is able to make theoretical evaluations about whether big data analytics can contribute to any of these areas. The second part reviews each area separately.

The third part of this analysis analyzes potentials of all big data techniques within big data analytics in relation to crisis communication & management on a more general level. This part ensures that this paper systematically consider all big data techniques. Through a utilization of previous analytical points from the prior parts, and the use of points and arguments made by Coombs, this part evaluates the potential of each technique. This provides insights into the theoretical possibilities and limitations of big data analytics in crisis communication & management.

# PART 1

## 4.1 - The capabilities of big data analytics

Big data analytics' ability to analyze patterns, trends, behavior and provides insights based on large amounts of complex data is an essential reason why this technology has caught the eye of academics and practitioners within a range of different fields and areas. Big data analytics expands the methodological possibilities to work with challenging amounts of data, and through this, provide new possibilities for working with complexity. Based on this point, it is possible to argue that the legitimacy of big data analytics lies in its ability to reduce complexity. Involving Luhmann's system theory as a theoretical perspective on complexity provides the basis to evaluate big data analytics as a complexity reduction tool. An assessment of big data analytics ability to work with complexity contributes with a valuable understanding about how big data analytics legitimize in general. This may help the paper to understand how big data analytics legitimize as a tool in crisis communication.

A main argument for why complexity reductions tools are interesting in the first place is that humans lack the necessary capabilities to reduce all of complexity in the world. The human ability to perceive complexity never reach the point where it can reduce all possible conditions and events of the world (Kneer and Nassehi, 1997, p.45). Any tool that can assist or improve complexity reduction processes stretch the human ability to analyze and understand complex situations.

### 4.1.1 - The use of big data analytics to reduce the complexity beyond systems

To learn more about how big data analytics work around complexity reduction, an initial question to ask is, if the use of big data analytics can defeat the challenges of environmental complexity on the behalf of human consciousness, or whether environmental complexity is too complex for big data analytics to handle. This is a leading question. The last option clearly seems to be the case because of the theoretical fact that environmental complexity goes beyond human comprehension (or consciousness) and is indescribable to humans (Luhmann, 1995, p.57).

Big data analytics only collect and analyze "comprehended information" such as words, numbers etc. to find correlations. In order for potential data in the environmental complexity to become comprehended information, systems must absorb this data and interpret it within their own medium. When systems absorb environmental complexity through this transformation process, it is no longer part of the environmental complexity, but the complexity of systems (Kneer and Nassehi, 1997, p.44). Because big data analytics is limited to comprehended information, it is also limited to the complexity within systems.

Another way to explain this is that human consciousness limits big data analytics since the methods are unable go beyond the dissemination media developed by humans (Thyssen, 2012, p.691). Consequently, big data analytics cannot find any universal truths or correlations outside the environment of systems. It is only possible to analyze the social constructions, interpretations etc., as environmental complexity is inaccessible (Luhmann, 1995, p.27). Big data analytics is not extraordinary in that aspect. It is rather an extension or improvement of humans' methods to analyze their "known" environment.

### 4.1.2 - Big data analytics within the boundaries of systems

As big data analytics cannot collect and analyze data beyond the boundaries of systems, the focus on complexity must change towards the complexity of systems. The central question becomes to what extend big data analytics can take on complexity challenges within the boundaries of both psychic and social systems. In order to do this, big data analytics must be able to collect and analyze data within the mediums of these meaning systems (Thyssen, 2012, p.696). As a reminder, psychic systems works in the medium of consciousness and social systems works in the language of communication (Thyssen, 2012, p.696).

The literary review of big data analytics found that with current methods it is only possible to collect and analyze numbers, text, audio and video material. This indicates that big data analytics can only work with social systems, and not psychic systems. All of these four data sources classifies as types of communication, and it is therefore legit to argue that big data analytics only is able to collect and analyze data within the medium of communication, and not consciousness. This may seem obvious to some, but it is imperative when considering the current capabilities and boundaries of big data analytics as a complexity reduction tool. It is imperative because it clarifies that big data analytics is only able to work with the complexity of social systems. This does not mean that psychic systems are irrelevant. Psychic systems and social systems intertwine and help each other reproduce, and create internal complexity (Thyssen, 2012, p.696). The point merely describes the fact that big data analytics cannot work with complexity outside social systems. What goes on in the psychic systems is invisible to big data analytics, and the abilities of the technology is limited to analysis of social interactions.

In continuation of this point, it is inevitable to examine to which extend big data analytics is able to analyze complexity of social systems. In connection to this, it seems relevant to ask how much social complexity big data analytics is able to capture. This question contributes with considerations about big data analytics' ability to produce analyses that reflects the overall picture of any situation and the social environment surrounding it. Any analysis within the boundaries of social systems requires the collection and analysis of communication. Big data analytics is therefore reliant on the access to communication. The more this access improves, the better are the conditions for big data analytics to comprehend the complexity of social systems. In order to gain access to communication, the technology need to be able to transform communication into data. The fact that big data analytics is able to collect communication in the forms of numbers, text, audio and video, means that the techniques can currently pick up a wide range of different types of communication.

However, there are still certain issues connected to the access of communication. Communication disappears instantly if dissemination media does not preserve it (Thyssen, 2012, p.691-692). It is unconceivable to think that it is possible to preserve all the communication constantly taking place. This means that big data analytics is unable to detect 'the whole mass of communication' within social systems on all levels. Big data analytics can only gain access to the communication preserved by the dissemination media. The dissemination media works as the gatekeeper that allow big data analytics to analyze the complexity of social systems. The extent to which these dissemination media capture communication therefore determines the amount of social complexity big data analytics is able to sample and analyze. Furthermore, this also implies that the dissemination media are decisive in terms of where big data analytics can find patterns, trends etc. in social systems. For instance, it is easier to gain access to communication on social media than communication happening at parties or at dinner tables. This means that the dissemination media have power over the access to communication. Big data analytics is not able to analyze the total amount of communication in social systems, but any increase or growth in dissemination media contributes with a potentially increasing accessibility.

## 4.2 - A framing perspective on the capabilities of big data analytics

The following part covers some additional points about the capabilities of big data analytics. This part uses framing theory as a supplement to systems theory to provide a different angle on the relationship between big data analytics and communication. As implied in the theoretical review of Coombs' framework, the concept 'perception' is central in crisis communication as to explain the development of organizational crises. As the previous analytical section concluded, it is impossible to enter psychic systems with big data analytics, and thereby analyze stakeholder perceptions. This section looks at whether it is still possible for big data analytics to understand and analyze perceptions through frames in communication, or whether big data analytics is unable to perform such an analysis.

### 4.2.1 - Analyzing stakeholder perceptions through frames

Since it is impossible to enter psychic systems, it is interesting to see if there are others ways to understand stakeholder perceptions. If crisis practitioners want to investigate the perceptions of stakeholders, they can only use communication and make interpretations of the perceptions among stakeholders. Though an interpretation of perceptions may not provide the perfect picture, it is still valuable knowledge in crisis communication & management. The big question is whether big data analytics can assist in the creation of this interpretation of perceptions.

A way to approach this question is to see whether big data analytics can analyze frames in the social context. The framing concept is interesting because it verbalizes some communication dynamics that shapes the social discourse and affect stakeholder perceptions (Entman, 1993, p.52). If big data analytics can understand the frames integrated in communication, it might be able to understand how the social discourses are changing stakeholder perceptions, and through this depict the current and future "perception tendencies".

Framing theory suggest that it is possible to spot frames in communication through certain keywords, stereotyped images and the application of certain logics (Entman, 1993, p. 52). The theory also states that it is impossible for humans to detect every frame in communication (Entman, 1993, p. 53). As the human mind is limited in terms of discovering frames, it is only more interesting to see if big data analytics can spot keywords, logics or stereotyped images in communication more sufficiently than humans can.

Research shows that big data analytics is unable to understand cultural behavior, traditions, logics, rhetoric etc. (Cambria and White, 2014, p.49). Academics from the field currently regard this issue as the biggest barrier within text analysis (Cambria and White, 2014, p.49). This barrier certainly limits big data analytics' ability to provide crisis practitioners with detailed analyses about social frames. The fact that big data analytics is unable to understand cultural behavior, traditions, logics and rhetoric, underlines the main difference between the human brain and big data analytics. Big data is not "intelligent" the way humans are, and it does not have the ability to perceive the environment as humans have. On its own, it is impossible for big data analytics to understand frames, and through this interpret stakeholder perceptions. It is up to the human brain to understand and interpret the consequences of the frames, as big data analytics does not understand the communication it analyzes.

However, big data analytics still have certain capabilities at its disposal. As the literary review of the big data field display, big data analytics can currently analyze dominant topics and words, analyze sentiments, identify dominant authors, and summarize text and importance sentences. The fact that big data analytics can analyze e.g. dominant topics show that it has the potential to identify frames. Therefore, big data analytics might still be able to assist in the process of interpreting stakeholder perceptions as an identification tool. The current capabilities of big data analytics improve the conditions to scan for social alteration in the environment. Analyses of dominant topics and sentiments might indicate what practitioners need to pay attention to, while the identification of dominant authors and spokespersons might provide knowledge about important actors that can influence the frames.

## 4.3 - Summary of part one

The previous analytical points circumscribe big data analytics as a complexity reduction tool that only is able to analyze social systems to the extent allowed by the dissemination media. Big data analytics cannot analyze all complexity within systems. Not even within social systems. When big data analytics analyze communication, it is actually unable to understand several dynamics in the communication. Examples of these dynamics are for instance logics, rhetoric and cultural perspectives. Big data analytics does not understand the implicitness in the language as humans do. Actually, it does not understand the language or the meaning behind the language at all.

Big data analytics is not 'intelligent' in the same way humans are. Humans are able to interpret communication on a detailed level and understand implicit meanings. Big data analytics cannot do this. Instead, big data analytics reduce complexity through generalizing large amounts of data derived from social systems, and create outputs such as patterns, correlations, trends. The quality of big data analytics is that the human brain is incapable of analyzing the same volume of data and find these correlations. Big data analytics is not a tool that can interpret social discourses on behalf of the human brain, but it can provide outputs that can contribute to the creation of meaning about social dynamics. This is perhaps the main contribution of big data analytics as a complexity reduction tool.

# PART 2

This part of the analysis takes on the five areas found in the preliminary analysis of Coombs' framework. The theoretical elaboration of the five areas clarify why these areas are interesting and how they fit into the over-all framework. The methodology explains the preliminary analysis and the selection process behind the discovery of these areas. To refresh the memory, the five areas in this part of the analysis are expectation gaps, stakeholder salience, situation awareness, social media monitoring during a crisis and crisis evaluation.

The aim of this part is to analyze whether big data analytics can contribute with any technical improvements or new approaches compared to the current methods and tools within these five areas. The analytical approach to each the five areas builds on a general three-step method. First, the analysis determines the information-needs within each area. Second, the analysis considers what exact big data techniques that have a potential to assist a crisis practitioner with his information-needs. Third, the analysis compares potential techniques with traditional methods and tools provided by Coombs, and analyze strengths and weaknesses of the big data techniques as well as the traditional methods. Based on these comparisons, the paper makes a theoretical evaluation for each area concerning whether big data analytics can contribute. The focus is to make theoretical conclusions about the combination of crisis communication and big data analytics. This part does not analyze any practical implications of the results.

## 4.4 - Expectation gaps

This section focuses on whether big data analytics has any contributions in terms of detecting expectation gaps. The first sub-section makes a technical analysis of expectation gaps to determine the information needs, and consider how big data analytics might be able to approach these needs. This knowledge contributes to the subsequent sub-sections that analyze the potential contribution of some big data techniques found interesting in relation to expectation gaps. After this, the analysis looks at the current tools and methods in Coombs' framework, and evaluate how these tools contribute, and what their strengths and weaknesses are. At last, there is a comparative analysis of the existing methods in Coombs' framework and the big data techniques, as well as an evaluation of big data analytics contributions to expectation gaps.

## 4.4.1 - The information needs in expectation gaps

### 4.4.1.1 - A technical analysis of expectation gaps and information needs

As the theoretical elaboration explain, an expectation gap exists when there is a gap between stakeholder expectations and organizational actions. These gaps can be either a performance gap or a perception gap. Though organizational performance matters, it is perception that determines whether a gap exist. Even if the organization perform accordingly to expectations, a gap still exists if stakeholders perceive it differently. Getting a clear picture of stakeholders' perceptions or expectations is the ultimate information need. However, as the first part of the analysis concluded, an issue about perceptions is that it is impossible to measure psychic systems. It is therefore not possible to calculate or understand the exact perceptions. Instead, it might be possible to identify the presence of an expectation gap through measuring whether stakeholders to have positive or negative perceptions.

The first part of the analysis concludes that big data analytics only can make an interpretative analysis of psychic systems through available communication. In addition, for big data analytics to analyze if stakeholder perceptions are positive or negative, it can only use communication stored in the dissemination media. This limits the potential of big data analytics for identifying expectation gaps, since a lot "escape" the dissemination. This means that an analysis of expectation gaps may not be representative depending on the available communication. Crisis practitioners must also be aware about the fact that only social interactions can provide information on perceptions or expectations. This means that not all information type of communication in dissemination works as suitable data. It also means that social media are ideal for extracting data, since the social media platforms contain a high amount of social interaction. Moreover, social media has free access to communication between stakeholders, which makes the platform ideal.

Another thing to be aware about is, that the communication in the dissemination media might be outdated. As both social systems and psychic systems constantly evolves, big data analytics is technically never up-to-date with the latest "state of mind". Big data analytics always work retrospectively when it comes to perception. This might not matter much if the conduction of the analysis happens within reasonable time of the social interactions, but it is important to be aware about when the different observations took place.

### 4.4.1.2 - Going into details with how sentiment analysis calculate reputation

Coombs mentions one reputation monitoring technique that might be interesting in relation to expectation gaps, since it is also a big data technique. This technique is sentiment analysis. Sentiment analysis as a way to measure if sentiments of stakeholders are positive or negative (Coombs, 2015, p.46+58). However, Coombs' description of sentiment analysis is limited, and he does not imply that crisis practitioners can use this technique to measure expectation. Still, the technique is interesting for its ability to measure positive and negative stakeholder sentiments and the following section therefore look at how sentiment analysis might be suitable for the calculation of expectation gaps.

Sentiment analysis is a big data technique that analyze whether a statement is positive or negative (Liu, 2012, p.6). It is important to point out that sentiment analysis does not analyze the underlying expectation of the utterance or the underlying perception behind the utterance. The most advanced technology within sentiment analysis is Natural Language Processing, and at this point in time it is not able to understand the meaning of a sentence (Cambria and White, 2014 p.49). The algorithms functions as a parrot. It is able to replicate interactions, but not able to understand the meaning behind them (Cambria and White, 2014 p.49). The only thing that sentiment analysis does is to calculate if stakeholder statements are positive or negative.

Colleoni et al. (2013) investigates the possibility of calculating online reputation of organizations using sentiment analysis. In their research they define reputation as the overall opinion about an organization by the stakeholders (Colleoni et al., 2013, p.317). This definition goes well with Coombs definition that reputation is an evaluation stakeholders make about the organization (Coombs, 2015, p.34). Colleoni et al argues that it is possible to use general feelings and sentiments on social media to describe the online reputation of an organization (Colleoni et al., 2013, p.317). Sentiment analysis absorbs subjective statements and interactions between stakeholders, and produce an objective value for online reputation of measurable standards (Colleoni et al., 2013, p.317). Hence, online communication offer a solution to monitor the evolution of an organization's online reputation over time (Colleoni et al., 2013, p.317).

Colleoni et al. base their calculations on whether the sentiment of the stakeholders is positive or negative. Earlier in the analysis, the paper concluded that an expectation gap exists if there is a negative or positive perception of an organization. By using the same calculations as Colleoni et al, this paper argues that it is possible to identify expectation gaps through sentiment analysis. However, the practical implication of using this method is still unclear.

### 4.4.1.3 - Aspect analysis to categorize sentiments

The way Colleoni et al. (2013) apply sentiment analysis in their research paper, could make it difficult to identify expectation gaps. This is because they evaluate the sum of all sentiments as part of calculating online reputations of organizations. Different types of stakeholders have individual expectations towards the organization (Coombs, 2015, p.55). As a result, stakeholders will most likely have different expectation gaps towards an organization. Some might have a positive expectation gap while others have a negative. There is a risk that the expectation gaps might level each other out, or a large amount of positive sentiments eclipses the few negative sentiments. The expectation gaps will then be difficult to locate and go by unnoticed. By measuring the overall sentiment, it is difficult to distinguish between expectation gaps.

Due to this problem, there is a need for the crisis practitioner to be able to distinguish between expectation gaps. For this purpose, a combination of aspect analysis and sentiment analysis might be valuable. Aspect analysis seek to examine the specific features of an entity a communicator finds positive or negative (Liu, 2012, p.59). Through, the application of aspect analysis, it is possible to carry out a categorization of the stakeholders' sentiments.

The algorithms of aspect sentiment analysis aim to identify aspects in a sentence, and hereafter recognize the sentiments associated with each of these aspects (Liu, 2012, p.59). If a crisis practitioner use aspect analysis, he should get a better overview of what aspects the sentiments are referring to about the organization. This might provide a method for the crisis practitioner to distinguish between expectation gaps and locate the critical ones.

### 4.4.1.4 - Limitations of sentiment analysis when identifying expectation gaps

Until this point, the analysis argues that sentiment analysis might be capable of identifying the existence of expectation gaps. However, an investigation about the limitations of sentiment analysis in relation to expectation gaps is also necessary. The following sections try to identify and analyze some of the limitations and challenges of using sentiment analysis for discovering expectation gaps.

So far, the analysis emphasizes the possibility of using sentiment analysis to measure an organization's reputation through online sources. The sources that provide the required data is vital in order to activate the use of sentiment analysis. Sentiment analysis cannot discover expectation gaps without accessible data. Thus, if an expectation gap emerges among e.g. shareholders, it might be difficult to detect it online or through other accessible data. That does not change the fact that shareholders are important stakeholders, and a crisis could evolve from a shareholder expectation gap. Online sources (or other sources valuable for sentiment analysis) does not necessarily guarantee the discovery of all expectation gaps between stakeholder groups.

A relating issue is that it is also difficult to classify and distinguish between different stakeholders on online & social media. It is therefore difficult to reveal the proportional importance of these stakeholders in relation to the organization. Still, crisis practitioners should be aware that not all stakeholders necessarily express themselves on social or online media about every concern or expectation. The fact that big data analytics is not able to go beyond the data sources and flow of communication means that the technology does not give any guarantees in terms of locating all possible topics stakeholders might discuss. If for instance a company sell a new product, it is first possible to tell if the product has created an expectation gap or not when consumers express their feelings about the product in an accessible form of communication. It is not possible to analyze the course of action or interaction made of about the product before and after this point in time.

When preventing a reputational crisis, it is essential to identify potential crisis in an early phase before it does harm to the organization (Coombs, 2015, p.44). It is not yet clear at what stage it is possible to identify a potential crisis by the support of big data analytics. With big data analytics, a crisis practitioner might be able to analyze the online communication flow to find expectation gaps. However, the analysis is limited to what the stakeholders are communicating. By analyzing the communication about an organization, it is impossible to predict expectation gaps before they occur. The crisis practitioner cannot evaluate how the stakeholder will react to the organizational actions until the organization carry them out. This means that sentiment analysis works in hindsight. The use of big data analytics might identify expectation gaps in an early stage, and allow practitioners to manage them before they develop into a potential crisis.

### 4.4.2 - The tools and methods in the crisis communication framework

The next step in this analysis of expectation gaps is to evaluate the current methods and tools suggested by Coombs related to information extraction and analysis. Coombs propose a range of different data gathering methods suitable for examining the existence of expectation gaps (Coombs, 2015, p.52). These methods are surveys, focus groups, interviews, use of informal contacts and content analysis (Coombs, 2015, p.53). The following subsections evaluate the contributions of the existing methods available.

### 4.4.2.1 - Benefits and limitations of using surveys

According to Coombs, surveys are useful for collecting information about the opinions, perceptions and attitudes of the stakeholders (Coombs, 2015, p.53). As surveys belong to the area of quantitative study methods, the purpose behind this method is to seek variables and correlations for statistical purposes, and obtain objectivity (Neuman, 2000, p.154). Though, surveys can include questions that require written responses, the survey approach is still rather specific and will not include the same opportunities to deviate from the original field of inquiry (Neuman, 2000, p. 16). Surveys follow a restricted path with fixed steps in sequence (Neuman, 2000, p. 124). Compared to qualitative methods, surveys are able to reach a larger amount of the stakeholders, because they are easier to distribute (Neuman, 2000, p. 16). This provide crisis practitioners with a better overall picture of stakeholder perceptions. In terms of expectation gaps, this means that the crisis practitioner better can track dissatisfaction in the general stakeholder environment.

Through the application of surveys as a method, it is possible to control the field of examination and ask questions of interest (Neuman, 2000, p. 124). In this regard, the crisis practitioner is able to direct the examination where he may suspect an expectation gap. However, the crisis practitioner is often unaware of potential expectation gaps, which might be problematic when using surveys for information gathering. The choice of method forces the crisis practitioner to direct the examination, while he does not necessarily know which questions to ask. Consequently, the crisis practitioner might miss expectation gaps because he did not ask the right questions.

A weakness of using surveys is that they are limited in terms providing into depth insights about perception compared to other methods (Neuman, 2000, p. 16). While surveys have a wide variety of questions and reach many different stakeholders, they do not provide a deep understanding of the perceptions (Neuman, 2000, p. 16). The right way to apply surveys in relation to expectation gaps might be to see them as a scanning tool that track dissatisfactions in the environment. Thereafter, the crisis practitioner can make further analysis about the reason behind the gap.

### 4.4.2.2 - Benefits and limitations of using interviews and focus groups

Coombs also highlight some qualitative methods to gather information about expectation gaps. The methods he mentions are interviews and focus groups (Coombs, 2015, p.53). Interviews and focus group have some similarities in their ability to gather deeper insights (Neuman, 2000, p.154). They focus on the interpretation and meaning behind a certain behavior and attempts to understand interactive processes and events (Neuman, 2000, p.154). The use of interviews and focus groups allow crisis practitioners to communicate directly with stakeholders, and ask question about topics of interest. This helps crisis practitioners to gain an "into depth" understanding about how stakeholder perceive the organization and the situation or topics discussed.

The interviewer controls the flow of communication and has the option to explore new directions that are suitable dependent on the findings (Neuman, 2000, p.154). Interviews are different from focus groups since they allow more focus on single participants, and isolate the participant from the influence of other stakeholders sitting in the same room (Neuman, 2000, p.154). However, a strength behind using focus groups is that the crisis practitioner may experience a more natural and less directed flow of communicate between the stakeholders (Neuman, 2000, p.154). In this way, the crisis practitioner get a better experience of how the stakeholders communicate about certain topics between themselves, and this may highlight some new aspects about the organization (Neuman, 2000, p.154).

Coombs suggest that crisis practitioners use key contacts for interviews that provide expert knowledge about the subjects that are up for examination (Coombs, 2015, p.52). Key contacts provide the kind of information that is not available when asking questions out in public (Coombs, 2015, p.52). Focus groups are particular useful when crisis practitioners wish to know more about specific stakeholders of interest. Compared to the use of surveys, the crisis practitioner gets improved conditions for focusing on specific stakeholder groups when using focus groups. However, making interviews and focus groups is a very time-consuming task, and it is not possible to include many stakeholders and a broad understanding of the stakeholder environment (Neuman, 2000, p.154). In this regard, the use of focus groups and interviews is not suitable for a broad scanning of the environment to find expectation gaps. Interviews and focus groups are a better choice of method, when the crisis practitioner know where to look for expectation gaps or has specific key stakeholders in mind with valuable information.

#### 4.4.2.3 - Benefits and limitations of applying content analysis

The last method that Coombs suggests as a tool to the identification of expectation gaps is content analysis (Coombs 2015, p.52). Compared to the other methods, content analysis is an analytical tool and not an information collection tool, and it involves analyzing the content of texts (Neuman, 2000, p.292). Content analysis provides a structure to approach the data produced from focus groups and interviews conducted by crisis practitioners (Coombs 2015, p.52). Hence, content analysis is "a next step" in the analysis process in extension to the interview and focus group processes. However, content analysis is not only limited as a tool that provide analytical insights of transcripts derived from interviews or focus groups. The content analysis method also apply for analysis of news stories and other written material (Coombs, 2015, p.52).

Content analysis structure and clarify the content in a text, but it cannot the content or meaning of the text (Neuman, 2000, p.293). Content analysis provides a great tool for crisis practitioners to gain an overview about the flow of communication. In that regard, it is quite similar to aspect analysis from big data analytics. However, because content analysis is unable to interpret information it cannot determine the existence of expectation gaps. Content analysis is also unable to determine sentiments in content or changes in sentiment over time. In this regard, content analysis works better as a supplementary to gain an overview of the content of the communication.

## 4.4.3 - Comparative analysis of methods

In the first sections about expectation gaps, this paper identified sentiment analysis and aspect analysis as methods that might be able to discover expectation gaps. This part of the analysis compares these two big data techniques with the existing crisis communication methods just explained in the previous sub-sections. This comparison highlights the main differences between these methods in terms of their contribution to the expectation gap area. This leads to an evaluation of whether big data analytics on a theoretical level is able to contribute to the identification of expectation gaps.

### 4.4.3.1 - Data sources

The traditional methods that Coombs suggest for discovering expectation gaps, selects samples from the stakeholder groups to gather information. This creates a risk that some expectation gaps go unnoticed, because the majority of stakeholders does not share the same perspectives or sentiments with a specific stakeholder or stakeholder group. Generally, big data analytics seeks to analyze large quantities of data. In theory, big data analytics provide the crisis practitioner with the potential to reach a larger number of stakeholders than through e.g. surveys. If so, there is less of a chance that expectation gaps will go unnoticed. However, big data analytics gain access to communication through dissemination media. That means that the access to dissemination media and content of these are the variables that determines whether big data analytics actually can reach and analyze a larger amount of stakeholder sentiments. Moreover, big data analytics can only analyze sentiments around topics discussed on e.g. social media. It is not possible to detect gaps outside the dissemination media.

Content analysis analyze data from dissemination media in a similar manner. However, (as stated earlier) content analysis is not able to discover expectation gaps itself, as this requires additional methods. Surveys, interviews and focus groups collects data by interacting with the stakeholders. This means that the organization influence the data in one way or another through interaction. Big data analytics provide methods to analyze data not generated through interactions of the organization, but through interactions between stakeholders only.

### 4.4.3.2 - Direct and indirect communication

The type of communication that is available for big data analysis is normally indirect, which entails that the communication is not directed to the organization, but rather from a stakeholder to other stakeholders. Using indirect communication means that the crisis manager is able to analyze the empirical field without affecting the flow of communication. Thereby achieving a more natural picture of how stakeholders perceive and communicate about the organization (Coombs, 2015, p.35). The only traditional method that comes close to achieving the same kind of indirect communication is focus groups. The use of focus groups involves both indirect and direct communication, because the stakeholders are between themselves and to the organization. The other traditional methods only involve direct communication. In this regard, big data analytics analyze a different kind of communication, which might bring some new insights. The disadvantage of only applying big data techniques is that the crisis practitioner cannot ask questions to understand the meaning behind the statements or examine some opinions in detail. This is only possible through qualitative methods such as interviews and focus groups, where the communication is directed towards the interviewer and thus the organization (Neuman, 2000, p. 154).

### 4.4.3.3 - Selection of stakeholder groups

It is difficult to select and examine a specific stakeholder group with big data analytics, when analyzing the communication on e.g. social media. On social media, different stakeholders are indistinguishable, unless the crisis practitioner chose to examine a specific community. In many cases, the consumers constitute the majority of the stakeholders on social media. If this is the case, then big data analytics represents the consumers' opinions and perceptions.

The user base on the social media does not represent all stakeholder groups in accordance to their importance. By using big data analytics, some stakeholder groups might be overrepresented as it is impossible to target specific stakeholder groups. With the traditional methods, it is possible to create a representative population of the stakeholder groups.

#### 4.4.3.4 - Time-frame

One of the big advantages when using big data analytics is that analyses can be done in a relatively quickly manner. By applying real-time analytics, the computer analyzes the communication directly from the dissemination media. If a machine is set to analyze the incoming data a specific way, it requires little effort for the crisis practitioner to do the analysis. This makes it easy to spot overall trends, because computers analyze the data is continuously. In crisis communication, it is essential to react fast to potential threats. These threats are impossible to spot fast with traditional methods, due to the time and resources it takes to conduct the analyses. This makes big data analytics a much more versatile tool to collect information.

### 4.4.4 - Contributions of big data analytics for expectation gaps

The previous sub-sections highlight some of the strengths and weaknesses of using big data analytics to identify expectation gaps compared with the traditional crisis communication methods in Coombs framework. Based on this comparison, the following part evaluate whether big data analytics has any contributions to the expectation area.

Based on the analysis, this paper suggests that sentiment analysis has a theoretical potential to discover the existence of expectation gaps among stakeholders on social media. In addition, this paper also suggest that aspect analysis is able to specify and locate where the expectation gaps are. The combination of sentiment and aspect analysis as a tool to identify and locate expectation gaps is interesting, and this paper suggest further empirical research on this area.

In terms of strengths and weaknesses of big data analytics compared to the existing methods in crisis communication, this analysis concludes different points. One central point is that big data analytics can contribute with a tool that provide crisis practitioners with a method to analyze the indirect communication on e.g. social media. Content analysis can also do this. However, as the analysis discovered, content analysis cannot identify expectation gaps alone, and this type of analysis is better suited to create an overview of the content in the communication.

Another point is that big data analytics as a method cannot control the flow of communication. As it is only able to analyze indirect communication through the dissemination media, it can only analyze what is “out there” and what is accessible. It cannot analyze topics or expectations around cases not present in these media. It is also unable to ask clarifying questions to detect why expectation gaps exist. Furthermore, it is also difficult for big data analytics to distinguish between stakeholder groups online.

However, an important advantage of big data analytics compared to the other methods is real-time analytics. Real-time analytics provide the crisis practitioner with the option of constantly analyze new communication. Big data analytics is the only method to provide such a tool. This can be important since the social environment can change rapidly, and it is important in crisis prevention to intercept when changes in the stakeholder "perception-environment" occurs.

# 4.5 - Stakeholder salience

As the theoretical introduction of Coombs' framework clarify, the assessment of stakeholder salience builds on to the analysis of expectation gaps to evaluate the potential of an expectation gap turning into a reputational crisis. The following section look into whether big data analytics can contribute to this assessment. However, the difficulty of analyzing this area is that Coombs does not provide any specific methods or tools to determine the power, legitimacy or willingness of stakeholders. The calculation of stakeholder salience seems rather subjective in method. It is therefore not possible to compare big data analytics with the current methods and tools for stakeholder salience in crisis communication & management. Nonetheless, there are still information needs to cover in this area, and it is therefore interesting to see if there are any potential contributions in the big data field. The information needs in stakeholder salience is the power, legitimacy and willingness of stakeholders. The following sections analyze whether any big data analytic tools have a theoretical potential to estimate or provide information that improve the measuring of power, legitimacy and willingness, and through this determine big data analytics capabilities of analyzing stakeholder salience.

## 4.5.1 - Using big data analytics to measure power

When going through the techniques from the big data field, only one technique seems interesting in relation to the calculation of power. This technique is social influence detection. Social influence detection calculates a stakeholder's ability to influence other stakeholders (Bakshy et al., 2011, p.66). The ability of a stakeholder to influence other stakeholders' dependents on how many other stakeholders the specific stakeholder reach on average through his online communication (Bakshy et al., 2011, p.66).

This is interesting in regards to power, because if a stakeholder has strong social influence he is better able to frame and affect the public discourses (Entman, 1993, p.54). As framing theory suggest, how something is presented affects how the audience perceive and process that information (Entman, 1993, p.54). A stakeholder with much social influence can affect the perception of the audience on the social media, and thus form the perception of an organization negatively (Entman, 1993, p.54). This means that the stakeholder might have power to influence other stakeholders through a strong social media position. As Coombs highlight the influence of social media for reputation threats (Coombs, 2015, p.22) it is possible to argue that a stakeholder with influence on social media has more power.

However, many components determine power. Coombs mentions money, contacts, communication channels and coalition formations among stakeholders as indicators of stakeholder power (Coombs, 2015, p.56). There are many factors to analyze when it comes to power. Big data analytics does not have the capability to analyze any of the above-mentioned factors. Social influence detection therefore only contributes with a limited understanding about the power of stakeholders that it purely about the power to influence on online media.

An issue with social influence detection is that Coombs argue that it is statistically more likely that a stakeholder with average social influence is the cause of a viral message (Coombs, 2015, p.27). This means that it is not necessarily the most influential online stakeholders that deserves the attention. This is also a reminder that willingness is a major factor in the development of a reputation threat. Social influence detection is maybe most valuable if the organization already know that a stakeholder has a high willingness score and the organization therefore is interested in determining the potential influence or power of this particular stakeholder. In this regard, social influence detection is better to apply when assessing power of a specific stakeholder with high willingness, than as an identification tool of which stakeholders to keep an eye on.

### 4.5.2 - Using big data analytics to measure legitimacy and willingness

While there is a small potential for social influence detection to contribute to the assessment of power, it is not possible for big data analytics to analyze willingness. The willingness of an stakeholder describes how keen the stakeholder is to confront the organization (Coombs, 2015, p.57). For big data analytics to determine this, it would require a technique that could measure willingness in communication. This is far too difficult for big data analytics at its current state.

For the assessment of legitimacy, this is almost the same case. However, there is the possibility of a small big data analytics contribution through sentiment analysis. As legitimacy refers to the validation of stakeholder actions from other stakeholders, it is theoretically possible to use sentiment analysis as a tool to assess whether other stakeholders respond positively or negatively to a comment or a post made by a certain stakeholder. However, this requires the localization of this comment and that other stakeholder already have responded. Crisis practitioners could use this method to calculate the legitimacy of e.g. a Facebook post with hundreds or thousands of comments to understand a debate, but at this stage a crisis practitioner probably already have received indications of the feelings in the stakeholder environment.

### 4.5.3 - Contributions of big data analytics for stakeholder salience

Based on previous analytical points, this paper argues that it is possible to use social influence detection as an assisting tool to evaluate the power of a stakeholder on social media. However, the results can only contribute to a broader perspective on stakeholder power. It is not possible with big data to calculate the exact power of a stakeholder. Social influence detect is not useful as a scanning tool either. It is most useful to use after the identification of stakeholders with a high willingness score.

The analysis also found very little indications that big data analytics is useful for determining legitimacy. The only interesting big data technique is sentiment analysis, and the potential of sentiment analysis is quite small. For the calculations of willingness, big data analytics is not able to contribute. The results in this part of the analysis shows that it is not possible to determine impact or likelihood scores for reputation threats through big data analytics. Thus, big data analytics cannot determine stakeholder salience without the help of human interpretation.

The benefits of using big data analysis to calculate stakeholder salience are questionable. The evaluating the power of a stakeholder is a relatively simple task and there is no guarantee that big data analytics do it better than human evaluation. The boundaries of social influence detection limit its use to social media influence, which is a small part of what power entails.

## 4.6 - Situation awareness

The theoretical elaboration of the framework clarifies that situation awareness involves three different areas. The perceptions of a situation, the comprehension of the situation and the ability to project future states. This part of the analysis first looks at the information needs in each of these areas and suggest different big data techniques that potentially can contribute to these information needs. After reviewing each of these areas, an analytical section then looks at the existing tools and methods in Coombs framework concerning situation awareness. After this, the last section concludes on the contributions of big data analytics in relation to the existing methods for situation awareness.

### 4.6.1 - The perception of the situation

The first part of the analysis concluded that it is impossible to understand stakeholder perceptions through big data analytics. Considering this, big data analytics cannot directly interpret the perceptions on behalf of the crisis practitioners. Crisis practitioners have to interpret perceptions themselves. However, it is still interesting whether any techniques can assist practitioners in gaining a faster or better understanding of situations. There are two techniques there are initially interesting. These are sentiment analysis and aspect analysis.

As previously described, sentiment analysis is able to determine negative and positive sentiments from communication. This might provide indications about the general perceptions of stakeholders. Aspect analysis, on the other hand, can contribute through analyzing dominant topics. This might help crisis practitioners with indications about how "big" the situation is in the eyes of stakeholders compared to other topics in the environment. The combination of these two techniques does not help crisis practitioners much in terms actual perceptions, but combined the two techniques are able to indicate the size and general feelings around the situation.

### 4.6.2 - The comprehension of the situation

Coombs highlights the need to comprehend the situation, when a crisis develops (Coombs, 2015, p.119). In a crisis, there is plenty of structured data to analyze, such as exposure of the media, demographic information, sales numbers etc. Tomaszewski et al. have shown how it is possible to use descriptive analytics, or more specifically geovisual analytics, in disaster situations (Tomaszewski et al., 2007). Geovisual analytics is a way to visualize geographical structured data in order to extract new insights (Tomaszewski et al., 2007, p.2). One way to apply geovisual analytics during a disaster, is to use visualize a map of the locations presented in the news media (Tomaszewski et al., 2007, p.4). Thereby, it is far easier for the crisis team to reach and plan their approach to provide help to the victims (Tomaszewski et al., 2007, p.5). However, this study focus on disasters, whereas this paper only examines reputation crises.

In the literature review of this paper, there was no literature on how to apply descriptive analytics to reputation crisis. Therefore, it is difficult to evaluate how to apply this technique on a theoretical level. However, it is the evaluation of this paper that it might be possible to outline certain trends of a situation through descriptive analytics. The use of descriptive analytics is very case-specific and depend highly on the situation. However, this paper recommends further empirical studies on how to apply descriptive analytics for reputational crises, since this technique might be able to determine multiple characteristics about the stakeholder that are communicating through social media. This could help the crisis practitioner to identify the unsatisfied stakeholder groups.

### 4.6.3 - The ability to project future states

The only current big data techniques that can project future states are techniques that use structured data such as predictive analytics. The big data techniques that use unstructured data are not currently capable of projecting future states, and there are no indications that it is possible in the near future. There are no indications that predictive analytics can project any future states in crises either. Predictive analytics can only use numbers to predict the weather or iPhone sales based on number of twitter messages (Gandomi and Haider, 2015, p. 143). As crises are perception-based, it is impossible for predictive analytics to make any calculations, and it seems that big data analytics cannot contribute within this area.

### 4.6.4 - Methods to create situation awareness in Coombs' framework

Coombs does not provide many tools for situation awareness. The two tools he suggests are crisis information logs and crisis knowledge maps. A crisis information log provides a system to keep track of the amount and movement of crisis-related information inside the organization. This helps crisis practitioners determine who know what and so on. This is important to keep practitioners organized and aware about the internal situation of the organization. However, crisis knowledge maps seem to be the key method to help practitioners gather data and create situation awareness. Crisis knowledge maps consists of the necessary contact information about stakeholders that help crisis practitioners to get in contact with the necessary stakeholders. Coombs highlight the importance of stakeholder relationships as an instrument to comprehend actual crises (Coombs, 2015, p.124). The data collection during a crisis situation should happen through stakeholder relationships (Coombs, 2015, p.119). It is important to create the stakeholder relations before a crisis occur and build relations with a variety of stakeholders (Coombs, 2015, p.38).

### 4.6.5 - Contributions of big data analytics for situation awareness

The contributions of big data analytics are weak in connection to situation awareness. The strongest potential contribution lies within determining the perceptions of a crisis. A consistent point in this analysis is that big data analytics cannot calculate perceptions. However, sentiment and aspect analysis could provide some indicators about the general feelings and the "size" of the situation in public. Compared to the existing methods, however, it is difficult to say how valuable this is for creating situation awareness in a real situation. Stakeholder relations is the main tool for creating situation awareness, and it seems that big data analytics cannot contribute with much in terms of the normal information needs in a crisis. However, there are some indications that descriptive analytics might be able to determine characteristics about the stakeholders communicating on social media. This could lead to further research on whether big data analytics could help crisis practitioners to profile the stakeholders interacting, and thereby provide overview of what type of stakeholder the organization is dealing with.

## 4.7 - Social media monitoring during a crisis

To this point, the analysis has already indicated that big data analytics is suitable for social media. This section looks at how big analytics can contribute to social media monitoring during a crisis. This is a small area in Coombs framework, since the focus of the framework lies more towards the classic crisis communication and management methods. However, Coombs highlights the importance of social media monitoring during crises, as well as the information needs related. It is therefore interesting to investigate the potential contributions of big data analytics. The first sub-section looks at the information needs and contributions of big data analytics in this area, while the second sub-section introduce the tools mentioned in Coombs framework. The last sub-section compares these and evaluate on the potential contributions of big data analytics in relation to social media monitoring during crises.

### 4.7.1 - Information needs and big data analytics' contributions

Coombs mentions three types of information needs in social media monitoring. These information needs relate to what stakeholders are saying, whether the comments are favorable or unfavorable and what the dominant topics are among stakeholders during a crisis (Coombs, 2015, p.157). A thorough investigation of the different big data techniques found three techniques interesting in regards to these information needs. The first two of these techniques are sentiment analysis and aspect analysis. These two techniques have already been found interesting within some of the other areas in this analysis. For social media monitoring during a crisis they can contribute to two of three information needs. Sentiment analysis can help crisis practitioners determine if comments are favorable or unfavorable and aspect analysis can determine what the dominant topics are. Together these two techniques can form aspect sentiment analysis, which means that practitioners actually can make analyses about the sentiment of specific topics.

The third technique that is found interesting is real time analytics. Real time analytics cannot cater to any of the three information needs, but it can combine with aspect sentiment analysis. The contribution of real time analytics is that it can activate the other techniques to make ongoing analyses of the information flow. Real time analytics measure the results directly, which means that it can provide crisis practitioners with fast analysis of the situation. The only information need big data analytics cannot cater to is what stakeholders are actually saying. The process leading up to this analysis considered descriptive analytics. However, descriptive analytics can only perform word counts.

### 4.7.2 - Current tools in the framework

For social media monitoring, Coombs suggest that crisis practitioners can use external crisis monitoring systems (Coombs, 2015, p.37+58+163+164). This includes systems that use big data analytics. Among these, Coombs also present sentiment analysis as a crisis communication tool (Coombs, 2015, p.58). However, Coombs does not go into depth with how to apply the technique or what it can do, which means that he is not very specific about how to work with social media monitoring during crises. As such, it is difficult to elaborate on how these current tools contribute, and what their strengths and weaknesses are for social media monitoring during a crisis from Coombs perspective.

### 4.7.3 - Contributions of big data analytics for social media monitoring during crises

It is difficult to compare the big data analytics techniques with Coombs suggestions for social media monitoring, since the framework does not have elaborations of the exact strengths and weaknesses of these tools for social media monitoring during crises. Sentiment analysis recur in the framework, but neither aspect nor real time analytics features among the external crisis monitoring systems. All three techniques have some contributions to social media monitoring during crises. Combined they can assist crisis practitioners with determining if sentiments are positive or negative towards the organization and what the dominant topics are. The three techniques can even determine the sentiments on a specific topic in real time. Compared to manual social media monitoring, big data analytics certainly have some contributions in terms of providing overview and fast analysis about some of these information needs. However, big data analytics cannot understand exactly what stakeholders are communicating. For this reason, manual social media monitoring still beats the technology on this specific information need.

## 4.8 - Crisis evaluation

The following part of the analysis consider the last of the five areas of information processing and analysis found in the preliminary analysis of the framework. This area is crisis evaluation. As mentioned in the theoretical elaboration of the framework, there are two types of crisis evaluations. These are the crisis management performance evaluation and the crisis impact evaluation. The first evaluation goes into detail with the performance of the organization and crisis management team, and the second evaluation goes into details with the damage or impact caused by the crisis (Coombs, 2015, p.164+166). This part examines if any big data techniques can contribute to any of these two crisis evaluations.

### 4.8.1 - Crisis management performance evaluation

The crisis management performance evaluation involves a carefully review of the crisis team's plan and their execution (Coombs, 2015, p.163). In this kind of assessment, the crisis team needs to question their actions during the crisis, examine potential mistakes and determine if the crisis team ignored any stakeholder queries (Coombs, 2015, p.163). To evaluate the performance of the crisis team, Coombs suggests examining the crisis records, stakeholder feedback, media coverage and word of mouth.

#### 4.8.1.2 - Crisis records

The crisis records are the first step to examine when evaluating the performance of the crisis team. Crisis records are internal data that the crisis team already are in possession of. It is difficult for this paper to evaluate whether big data analytics has any contributions when analyzing internal data. The reasoning is that internal data is very case specific and are not necessarily classifies as big data. In this regard, it is difficult for this paper to generalize any analytical points about the contribution of big data analytics.

#### 4.8.1.3 - Stakeholder feedback

To gain external feedback on the crisis team's performance during a crisis, it is important to listen to a wide variety of stakeholders. This creates a broad picture of the opinions of the organization's stakeholders. When applying big data analytics, it is not possible to select certain stakeholders, or decide what stakeholders to ask. As a result, it is difficult to target specific stakeholder groups, or make sure that the data presents stakeholder groups equally. If the crisis practitioner applies big data analytics to the social media, the customers are likely overrepresented.

The use of big data analytics does not provide the crisis practitioner with the option to direct the communication. In this regard, it is not possible to inquire about specific topics of interest. Big data analytics is limited to the available data. Furthermore, big data analytics is not able to extract "in depth" opinions and meaning based on stakeholder communication. It can determine the sentiment of the statements. However, it is difficult for the crisis practitioner to determine if the communication concerns the specific crisis. The crisis practitioner cannot identify important details about the stakeholders' opinions. This paper estimates that the contributions of big data analytics in this area are small.

### 4.8.1.4 - Media coverage

The media has great influence on how the public perceive a crisis (Coombs, 2015, p.166). Often, the media is the only source of information that reach the public (Coombs, 2015, p.166). In this regard, the media has great power to control and frame the information relating to a crisis (Coombs, 2015, p.166). Measuring the media coverage is therefore a good indicator of how well the crisis practitioners did in terms of managing with the crisis (Coombs, 2015, p.166-167).

There are three important aspects of media coverage to examine. The first is the percentage of media reports that use the messages of the crisis practitioner (Coombs, 2015, p.167). The second is a comparison between the organizational frame material and the counter frame material (Coombs, 2015, p.167). The third is the accuracy of the crisis-related information appearing in the media (Coombs, 2015, p.167).

The news is gradually gaining ground on the online media (Coombs, 2015, p.166). Thus, the digital dissemination media store the news, which makes it easier to analyze the media through big data analytics. The first of the three media coverage aspects involves identifying statements and spokespersons in the news flow. Pouliquen et al., have developed a big data analytics software to detect quotations and spokespersons in multilingual news media (Pouliquen et al., 2007). The software uses the big data technique of information extraction. Information extraction first identify the relevant news articles. Then, it examines the articles to identify and extract certain information. By the help of natural language processing, it is able to identify citations and recognize the spokesperson. This technique has the potential to assist the crisis practitioner by determining the percentage of media reports that use messages from the crisis practitioner.

It is far more difficult to apply big data analytics to the second and the third aspect of media coverage. These aspects require more in depth understanding of the content in the articles. Big data analytic is not capable of analyzing text to such an extent to identify the accuracy of the information or compare the frame material. As a result, big data analytics does not have potential to assist the crisis practitioner with the two last aspects.

Currently, much of the digital news media broadcast through audio or video. The transcript-based approach in audio analytics transcribes the data after which text analytical techniques apply. If the information extraction technique has potential to assist the crisis practitioner, then the transcript-based approach expands the possible data sources to include audio data.

#### 4.8.1.5 - Word of mouth

In order to explore the word of mouth, Coombs suggests to collect data on the social media as it is easy to access (Coombs, 2015, p.167). Since it is essential knowing how the communication flourish, he suggest to divide the statements between the crisis team (Coombs, 2015,p.167). At the post-crisis stage, time is not a critical aspect compared to the crisis stage (ref). In this regard, there is more time to read the communication more thoroughly and examine how the stakeholders perceive the organization. For this reason, it is not very relevant to apply big data analytics, since it is not yet possible to analyze communication in depth.

### 4.8.2 - Crisis impact evaluation

The second evaluation the crisis team should conduct is an estimation of the crisis impact on the organization. According to Coombs, "*damage assessments provide a tangible indicator of crisis management success or failure*" (Coombs, 2015, p.162). Coombs is not very clear in his presentation of useful information about the impact evaluation for the crisis practitioner. The criteria of success depend on the organization (Coombs, 2015, p.166). This analysis therefore goes in depth with the performance measurements that Coombs highlights. These are the financial performance of the organization and reputation impact. (Coombs, 2015, p.166-167).

#### 4.8.2.1 - Financial performance

The financial performance of the organization involves certain financial ratios such as sales numbers, market shares and stock prizes (Coombs, 2015, p.167). Coombs does not describe these financial ratios in depths, which makes it difficult to determine if big data analytics have the potential to assist the crisis practitioner. Furthermore, it is not possible to determine if the organization has big data available to calculate the key figures based on Coombs framework. The data might be structural data and come in large volume. However, if the variety of the data is usually low, it makes the data relatively easy to handle. In this case, the data does not inquire ‘cost-efficient innovative forms of information processing´ to handle the data. Thus, this paper does not know if the calculation of financial ratios involves big data. This paper does not go more in depth with the evaluation on the use of big data analytic to calculate financial performance.

### 4.8.2.2 - Evaluating the implication on reputation

When evaluating the crisis’ implications on the reputation, the best indicator is comparing the reputation of pre and post crisis (Coombs, 2015, p.166). It is important to measure the reputation in both traditional and digital media (Coombs, 2015, p.167). The traditional ways conduct the information gathering through interviews, surveys or focus groups (Coombs, 2015, p.163). The limitation of these methods allows them only to determine the reputation at the current time. Thus, it is not possible to calculate the reputation of the past (Coombs, 2015, p.166). However, if the organization track its reputation on a regular basis, it is possible to measure the crisis’ impact on the reputation (Coombs, 2015, p.166).

Due to the importance of internet communication, the crisis management team should also collect data on online media (Coombs, 2015, p.166-167). However, Coombs argue that online stakeholder evaluations are an imprecise substitute for actual reputation measures (Coombs, 2015, p.168). Though, it is a useful tool when direct reputational measuring techniques lack (Coombs, 2015, p.168). In order to include all stakeholders, it is important to conduct reputation measure through both traditional and digital media (Coombs, 2015, p.167). Coombs suggests applying sentiment analysis to measure the reputation of an organization but does not go into depth on how to apply big data (Coombs, 2015, p.58). One of the big advantages of using sentiment analysis to measure the reputation is that it is possible to do at any point in time. Even after the crisis, it is possible to collect data to calculate the reputation for pre and mid crisis. This makes it a far more dynamic tool than traditional methods.

According to Coombs, specificity is key when evaluating the crisis team’s performance and a good or poor evaluation is too general (Coombs, 2015, p.164). In this regard, sentiment analysis is not very specific when assisting the crisis practitioner. However, Coombs does not seem to be knowledgeable about aspect analysis, which is able to categorize the data and sort feedback into groupings. This tool has the potential to make it easier for the crisis team to gain an overview of the topics discussed and determine the sentiment of each topic.

### 4.8.3 - Contributions of big data techniques for crisis evaluation

In this part of the analysis, the paper examined which big data techniques that have potential to assist the crisis manager at the post-crisis stage. This involves both evaluations of the crisis team's performance and measurements of the impact of the crisis. At the performance evaluation stage, the analysis was not able to analyze the contributions of big data analytics, when analyzing the internal crisis records. Internal data is very case-specific and it is difficult for a theoretical study to generalize any conclusions.

This analysis found little indication that big data analytics has the potential to contribute to analyzing the word of mouth and stakeholder feedback. The argument is that big data analytics is not able to understand meanings and opinions to an extent that is valuable for the crisis practitioner. When analyzing stakeholder feedback, big data analytics is not able to direct the communication of the stakeholders to get feedback on topics of interest.

For tracking the crisis practitioner's performance in media coverage, information extraction and transcript-based approach have the potential to provide assistance. These techniques have the potential to measure the percentage of online media reports that use the messages of the crisis practitioner. This concerns both text, audio and video-based media reports. Coombs does not provide any specific tools for this purpose. In this regard, these techniques have the potential to offer great value for the crisis practitioner, so he does not have to look through the news reports manually.

Tracking the crisis impact involves the measurement of organizational reputation and financial performance. According to Coombs, it is preferable to measure the impact on the reputation through both traditional and digital media. At the post-crisis stage, time pressure is not as great a factor. Therefore, traditional methods apply more easily because they require more time and resources to conduct. Coombs argue that the traditional methods are more precise because the crisis practitioner is able to reach all stakeholder groups and provide a more nuanced picture. However, it is valuable for the crisis practitioner to conduct both traditional and digital measurements. To help determine the crisis impact on the organization's reputation in digital media, sentiment analysis and aspect analysis have potential to assist the crisis practitioner. Furthermore, if the organization does not track their reputation on a regular basis, big data analytics might be the best way to measure the impact on reputation. This is because big data analytics have the potential to collect and analyze data from pre and post crisis at any current time.

This analysis did not find any use for big data analytics in the financial performance evaluations. Coombs description of financial ratios is too vague to determine if big data analytics might be useful. Since the analysis is theoretical, it is not possible to determine the practical use of big data techniques. Therefore, this paper recommends several empirical studies to determine the practical implications. First, a study on the use of sentiment analysis and aspect analysis to measure the impact on reputation post-crisis. Furthermore, a study on the use of information extraction and transcript-based approach in audio analytics to determine whether they are useful tools for the crisis practitioner to track the media coverage.

# Part 3

## 4.9 - A review of big data techniques

This part is the third and last part of the analysis. The first part of the analysis explored big data analytics capabilities as a complexity reduction tool and as a tool to analyze communication from a combined perspective of systems and framing theory. The second part of the analysis examined the contribution of big data analytics within five specific crisis communication & management areas that a preliminary analysis of Coombs framework found particular interesting. The two first parts bring valuable knowledge about the theoretical contributions of big data analytics in crisis communication & management. However, it is important for this analysis to express the considerations of the contributions behind all of the identified techniques. To express this process of analysis for every technique and compare it to each of the five areas in part two would be too repetitive and comprehensive to carry out. Nevertheless, there are still several uncovered techniques, as well as a need to express why these techniques have not been included until this point.

The following part of this analysis reviews all big data techniques and discuss their potential in assisting the crisis practitioner on a more general level. It is important to ensure that the paper methodically express considerations about all techniques. By using the previous analytical points and the framework of Coombs, this paper argues for the potential of each technique. It offers an overview of the potential use of big data analytics within the crisis communication discipline. Furthermore, it provides the reader with insights into why some techniques cannot assist the crisis practitioner.

### 4.9.1 - Text analytics

Within the field of text analytics, the literary review of big data analytics found that there are four overall techniques. These are sentiment analytics, information extraction, question answering and text summarization. Throughout this analysis, there are several stages where sentiment and aspect analytics may assist the crisis practitioner. In fact, the technique is the most promising within big data analytics because of the value of knowing the stakeholders' sentiment towards the organization.

Information extraction gathers basic information from large number of articles. The technique is able to recognize authors, quotations, organization etc. This paper found the technique to have potential in tracking the media coverage at the post-crisis stage. This provides the crisis practitioner with a method to track how often the media cites the organization's spokesperson in comparison with another spokesperson. This offers a good indicator of how well the crisis practitioner did in covering the news flow.

The big data analytics technique ‘question answering’ is valuable when the analyst requires simple short answers. This paper did not find any need for this kind of analysis in Coombs’ framework. Generally, the valuable insights for the crisis practitioner involves thorough analyses of perceptions, opinions and interpretations. As a result, this paper does not recommend the use of question answering for the crisis practitioner.

The primary use of text summarization is to subtract and convey the key information in a text. This technique works well if the analyst has to gain an overview of the content in a large number of documents. However, these summaries frame the content differently and do not capture accurate meanings of the original article. It is important for the crisis practitioner to gain a detailed overview of how the media frame the situation (Coombs, 2015, p.111). In this regard, text summarization is not able to provide an accurate overview of how the media frame the situation. It is therefore questionable if text summarization is able to contribute to the crisis communication discipline.

### 4.9.2 - Audio analytics

There are two different overall techniques in audio analytics. These are transcript and phonetic based approaches. The phonetic based approach analyzes sounds and phonemes of speech. It does not analyze what a person says, but how the person articulates. In this regard, the technique does not capture opinions and the meaning in a statement. The application of this technology would be able to tell a crisis practitioner if a certain stakeholder is angry or happy. However, the reason behind the mood would be invisible, and the crisis practitioner would not know if a certain word or utterance connects negatively or positively to the organization. As the technology would not provide any insights into the stakeholder perception, the technology is not valuable enough for crisis practitioners.

The transcript-based approach transcribes spoken language into text. In order to extract meaning from this data, the analyst need to apply text-based techniques. In this transition process (from voice to text), there is a risk that meaning of an utterance gets lost. This especially concerns irony and/or feelings, because these are difficult to determine in a text format. Consequently, the advantage of listening to the utterance is lost, because there is no difference between the voice data the technique produces and data in text format. However, the technique allows text-based technology to expand into spoken language and analyze speech. As a result, there is more data for big data technology to analyze. This technique has particularly potential at the post-crisis stage, when the crisis practitioner wants to gain an overview of the media coverage. Many of the news media use audio and video channels, which is accessible to analyze by the help of the transcript-based approach.

### 4.9.3 - Video analytics

The development in video analytics has not progressed as far as other types of big data analytics (Gandomi and Haider, 2015, p.141). At this current stage, video analytics is able recognize people and identify movement on video (Regazzoni et al., 2010, p.16). The type of insights that video analytics generate is restricted to identification of changes and visual recognition (Regazzoni et al., 2010, p. 16). It is not yet possible to determine the content of a video based on video analytics (Regazzoni et al., 2010, p. 16). In this regard, video analytics cannot analyze the communication in the dissemination media. The information required by the crisis practitioner, is often deep insights on perceptions, interpretations and opinions, which is far too complex for video analytics. With video analytics, the analysts cannot comprehend the situation and by no means analyze how the video frame the situation or how the audience perceive it. This require human interpretation skills. As a result, the current technology is insufficient in contributing to anything within the crisis management framework. However, the paper does not exclude the use of video analytics in crisis communication if the techniques progress.

### 4.9.4 - Social media analytics

The social media provides many different types of data to analyze through big data analytics. Although techniques in social media analytics only relates to the connectivity between individuals (Gandomi and Haider, 2015, p. 142). It involves interactions between users on the social media. In the review of the five area in the framework, this paper only suggested social influence detection to have the potential to assist the crisis practitioner. Social influence detection has the potential to determine the online power of the stakeholder. However, online power on social media is only a small fraction of what power involves. In this regard, the technique is restricted and human evaluations are necessary to determine the power of the stakeholder. Therefore, this paper does not see the techniques to be very useful for the crisis practitioner.

According to Coombs, any stakeholder is able to create a potential threat to the organization (Coombs, 2015, p.27). Due to the complexity of social media analytics, it is a massive task to analyze all stakeholder on social media. This undermines the use of social media analytics before a potential threat arise. Social media analytics takes time to conduct and requires machine learning to complete the analysis. In this regard, the techniques are not very useful under time pressure. When a threat occurs, the crisis practitioner has to react quickly. As a result, social media analytics is neither very useful for the crisis practitioner in crisis recognition or during a crisis.

The field of social media analytics is very new and complex (Fisher et al., 2012, p. 50). It is difficult to outline the use of social media analytics and it is still unclear if social media analytics has something to contribute with in the field of crisis communication. Due to the uncertainties, it is difficult for this paper to analyze the use of social media analytics more in depth. Therefore, this paper cannot exclude its use in crisis management entirely even through it did not find the techniques useful for the crisis practitioner.

### 4.9.5 - Big data techniques for structural data

There are three overall areas in big data analytics that analyze structural data, which are predictive, descriptive and prescriptive analytics. The monitoring systems that Coombs describe, gathers structural data about the stakeholders on social media. However, Coombs does not go into detail on how the structural data from monitoring systems is valuable for the crisis practitioner.

An analysis based on data from social media involves other forms of communication such as Facebook likes, activity and behavioral patterns (Gandomi and Haider, 2015, p. 142). By applying descriptive analytics to the data from social media it is possible to gather information about the demographic of stakeholders and if they are regular or new visitors of the organization's social media profiles. This kind of information might help the crisis practitioner to specify, which stakeholder groups that have a negative perception of the stakeholder. However, it is difficult for descriptive analytic to capture deep interpretations, understandings and feelings. These limitations are a result of the structural categorizations, which has to be in the form of number, letters or single words (Gandomi and Haider, 2015, p. 143). Hence, it is difficult for structural data to capture the essence of a social phenomenon such as crises.

Based on Coombs framework, it is difficult to determine the value of descriptive analytics for the crisis practitioner. Coombs does not provide sufficient information on which insights that are valuable to the crisis practitioner. As a result, it is difficult to determine the theoretical value of descriptive analytics in crisis communication. The methodical choices on how to apply Coombs entails that it is problematic to conduct a theoretical examination of the use of descriptive analytics. Because of this, the paper does not go more in depth with descriptive analytics in the five information-gathering stages in Coombs framework.

The use of predictive and prescriptive analytics is far more limited in the field of crisis communication. The framing of a situation has a big influence on how the stakeholders interpret and perceive a situation. Furthermore, it requires a cultural understanding to determine the developments of a crisis. This entails that crises become very unpredictable for machines that do not share a cultural understanding and cannot understand the framing of a situation. Crises are case-specific in this regard, and this paper does not see much value in predictive analytics for the crisis practitioner.

Prescriptive analytics helps an organization to choose between different actions, by projecting the outcome of each actions. Since it is very difficult for a machine to project the future outcome of a crisis, these calculations are not very precise. As a result, this paper does not recommend the use of prescriptive analytics for the crisis practitioner.

# Chapter 5 - Discussion

The following section discuss the methodological and theoretical backbone of this paper, as well as what the results of this paper means to the crisis communication discipline. The main purpose is to provide some reflections on the possible strengths and weaknesses in this paper. This is in terms of how it is possible to criticize the approaches of this paper, and to argue for the method in this paper.

## 5.1 - Evaluation & critique of the method

An essential part of this discussion is to evaluate and question the structure of the analysis. Each part of the analysis has a different purpose and contribution to answering of the research question. However, how does each part contribute? – And are there any weaknesses in this three-part structure?

In terms of weaknesses, there are a couple of things to discuss. These things specifically concern part two of the analysis. A main purpose with part two is to provide the reader with an idea about what areas big data analytics might contribute through a presentation and examination of five specific areas in Coombs' framework. The main problem with this approach is that a large part of the analysis process leading up to the selection and analysis of these five areas is not included in the analysis. First, the analysis does not show the process of the preliminary analysis of Coombs framework. This makes the reader unable to judge whether these five areas are the only ones relevant in Coombs framework. Second, after the selection of these five areas, there has been a long analysis process of comparing every big data technique to each of the areas. The second part of the analysis does not display the argumentation for excluding certain techniques from the five specific areas. Therefore, it is difficult for the reader to judge if other big data techniques might work in a specific area, or if the exclusion of a certain technique is wrong. The third part of the analysis attempts to make up for these weaknesses through elaborating the thoughts about the potential of each big data technique in crisis communication. Still, it is difficult to include the whole process of analytical choices.

On the other hand, part two of the analysis does contribute with some of the most essential conclusions and arguments in terms of answering the research question. The strength about the methodological choices behind part two contribute with a simple and manageable structure. The principles of focusing on areas that includes gathering and processing of data, speaks to the capabilities of big data analytics. Therefore, the analysis avoids presenting a bunch of areas in Coombs' framework that are not relevant in any way possible. As such, the choice of making a preliminary analysis of Coombs framework benefits the paper in terms of focus. The literary review of big data analytics has the same contribution as it delimits the big data analytics field, and therefore focus specifically on the current processing and analysis techniques available.

Combined, all of the three parts in the analysis contribute to the research question. The contributions from part two and three specifically take on the techniques and compare them to the discipline of crisis communication. However, part one of the analysis also has some important contributions. Though it does not compare any big data techniques with the framework, it considers the general role of big data analytics in relation to communication, perception and complexity issues. Part one contribute throughout part two and three, because it creates an understanding about the general capabilities of big data analytics, which leads to further considerations when analyzing what big data analytics actually can do (or not do) in relation to specific issues and areas in crisis communication and management. Especially the contributions of understanding the relationship between perception and big data analytics is important throughout the analysis.

Another thing important to discuss is the theoretical approach of this paper. There is a clear weakness with this approach concerning the fact that this paper cannot produce any insights into to practical implications of using big data techniques in crisis communication and management. The analysis focus on whether it is thinkable that the techniques can solve any information needs or theoretical issues, but it is impossible to say something how the application of the techniques within the five areas actually would work out. For this reason, this paper recommends further practical research every match found between big data analytics and the crisis communication discipline. However, the theoretical approach still seems to be the best choice for this paper, as the purpose is to explore the potential contributions of big data analytics within every area in the crisis communication discipline. A case study would not be able to contribute with the same broad perspective on crisis communication, although a case study might bring insides about the practical implications of big data analytics.

### 5.1.1 - The theoretical choice of Coombs

The theoretical choice of Coombs framework supplies a fundamental structure that is important for carrying out a theoretical investigation. Without the framework, a theoretical investigation would require an entire literary review of the crisis communication discipline to establish the current methods, tools etc. available. However, there are some implications of using this framework. First, the framework does not go into depth with single areas on the same level as articles only treating a single subject. In some areas of the framework, Coombs does not elaborate in details how to perform certain analyses systematically. Other articles could have provided more details about specific areas. This means that it at times is harder to tell if big data analytics actually can contribute. Second, it is possible to question whether one academic can represent the whole discipline. For this reason, it is an essential fact that Coombs is probably the most recognized scholar within the discipline. Still, Coombs is fundamental to this paper. The choice of Coombs does not only provide a structure. It also helps define the perspective on organizational crises as a phenomenon deriving from stakeholder perceptions. This puts an emphasis on the most important dynamics in crisis communication.

## 5.2 - Discussion of the results

The results of this paper are very theoretical. However, they still reveal some clear indications about the value big data analytics in crisis communication. The results seem to confirm that computers have not yet superseded the capacities of the human brain on several levels, and have a long way to go. Big data analytics can only do simple things with communication, and is still highly dependent on crisis practitioners to make the interpretations and create meaning of the data. In relation to this, an interesting question to discuss is what the results of the analysis might mean for the future job functions in crisis communication area. One thing seems certain. Big data analytics will not replace humans any time soon. However, big data analytics might still have a future in the crisis communication discipline. This is especially due to the increasing amount of data available through social media and other online media.

The results of the analysis show repeatedly that the contributions of big data analytics lie towards the online and social media areas. Many tasks within crisis communication & management does not necessarily involve these media, and it seems that big data analytics is unable to create a whole revolution within the crisis communication discipline. Qualitative methods and subjective situational awareness still seem to be a virtue and a cornerstone within crisis communication & management. However, as the results display, there are still reason to explore the potential of big data analytics within some areas, as an assisting tool to cope with large amounts of data.

# Chapter 6 - Conclusion

Big data is a relatively new field in academic studies, and the potential of big data analytics is still unknown within several areas. This study attempts to take on the theoretical possibilities and limitations of big data analytics in crisis communication & management, and it examines whether big data analytics has any contributions to the existing in the crisis communication discipline. The core of the methodological approach has been to compare the current big data processing techniques with the current methods and tools from the area of crisis communication to achieve the necessary insights. As a result, the paper has carried out an extensive analysis that elaborates on the potential of big data analytics within central areas in crisis communication theory, as well as examine the potential of every data processing technique derived from an extensive literary review of the big data field. As a supplement to this analysis, the paper has also included perspectives from framing and systems theory to consider the possibilities and limitations of big data analytics on a general level.

## 6.1 - The possibilities and limitations of big data analytics in crisis communication & management

The use of big data analytics provides individuals with a tool to reduce complexity. It expands the methodological possibilities of solving the challenges of working with large amounts of data. Big data analytics is, however, limited when analyzing communication, as it is not able to defeat the challenges of analyzing environmental complexity. Big data analytics is limited to analyzing quantifiable and comprehended communication such as of numbers, words, audio or video. Because big data analytics cannot analyze environmental complexity, it cannot achieve universal truths. It is rather an extension of the human capabilities of analyzing the world.

The current technology cannot analyze psychic systems directly, and nor will it probably ever become capable of it. However, big data analytics is able to analyze communication in social systems. The interactions of psychic and social systems may support an understanding of psychic systems through analyzing the communication in social systems, though it will never to a full extend be possible comprehend the human minds by analyzing interactions in society. An issue about communication is that a lot of it disappears instantly. Dissemination media solves a part of this issue. Dissemination media store the communication and make it possible to apply big data analytics. Without dissemination media, it is not possible to apply big data analytics.

Social systems get more complex as they reduce the complexity of the external environment. Big data analytics cannot analyze the external complexity, but can help individuals to understand the complexity within systems. It is able to create outputs such as patterns, correlations and trends based on large amount of data derived from social systems. The human mind is capable of making deep interpretations and understands implicit meanings. The quality of big data analytics is that it can generalize large amount of data and create outputs that complement the human capabilities.

## 6.2 - Analyzing meaning and interpretation in communication through framing

A premise in this study is that crises are perceptual and the shared the perception of the stakeholders determines if the organization is in a crisis. The conceptualization of crises as social constructed phenomena limits the potential use of big data analytics. The framing of information influences how the audience perceives the crisis. This makes it very difficult to predict how stakeholders' perceptions develop, because perceptions can change radically through framing. This is a big challenge in the information gathering process, which makes it relevant to examine if big data analytics is able to identify frames in the communication.

Big data analytics is unable to understand concepts such as cultural behavior, logics, traditions, stereotyped images and rhetoric. This barrier certainly limits big data analytics' ability to provide crisis practitioners with detailed analyses about social frames. On its own, it is impossible for big data analytics to understand frames, and through this interpret stakeholder perceptions. It is up to the human brain to understand and interpret the consequences of the frames, as big data analytics does not understand the communication it analyzes.

Nevertheless, there still is some potential for big data analytics to identify frames. The fact that big data analytics can analyze dominant topics in a text shows the potential. Therefore, big data analytics might still be able to assist in the process of interpreting stakeholder perceptions as an identification tool. Big data analytics has the potential to create an overview for the crisis practitioner, however, cannot interpret the frames itself.

## 6.3 - Potential contributions of big data analytics

The following section outlines the potential contributions of big data analytics within five selected areas of Coombs' framework. The potential contributions of big data analytics are the product of a theoretical assessment based on a literary review of the big data analytics field. The analysis selected the five areas by identifying areas within crisis communication & management, which involves collecting and processing of information. This is where big data analytics might have some contributions. Since the analysis is theoretical, it is not possible to evaluate the practical implications of using big data analytics to identify expectation gaps. Therefore, this paper suggests further empirical research wherever the analysis identify a potential contribution.

### 6.3.1 - Expectation gaps

The information gathering process of identifying expectation gaps is part of the crisis preventions stage. By identifying expectation gaps at an early stage, it is possible to prevent them turning into a crisis. Based on the analysis, this paper found great potential for sentiment analysis to discover existing stakeholder gaps among stakeholders on social media. To help locate and identify expectation gaps, the crisis practitioner can combine sentiment analysis with aspect analysis. Through this, the crisis practitioner can better locate the aspects of the organization where stakeholders have negative sentiments.

Coombs suggests several information gathering methods in his framework, which have multiple advantages and weaknesses compared to big data analytics. Primarily, big data analytics provides the crisis practitioner with an opportunity to analyze indirect communication in digital communication devices. If it is successful to implement the techniques, big data analytics can analyze the flow of communication when representatives from the organization are not present. This gives the crisis practitioner the opportunity to analyze communication without any disturbance or influence from the organization. However, there are also weaknesses of not being able to control the flow of communication. By analyzing indirect communication, it is not possible to steer the flow of communication to topics of interest. Furthermore, the crisis practitioner cannot ask the stakeholders about future actions and decisions of the organization. The crisis practitioner is limited to analyzing available data.

Digital dissemination media does not represent all stakeholders equally and therefore; it does not provide a representative overview of the stakeholder sentiments. Since big data techniques are not able to distinguish between stakeholders on digital media, some expectation gaps are difficult to spot.

Compared to existing methods, big data analytics has a big advantage in real time analytics. This provides the crisis practitioner with the option to analyze new communication continuously. Through this, the techniques might be able to identify new expectation gaps at a faster rate than existing methods if successfully implemented.

Overall, big data analytics provides the crisis practitioner with the opportunities to analyze communication that would otherwise not be possible to analyze through existing methods. In this regard, there is great potential to apply sentiment analysis and aspect analysis as a supplement to existing methods. Therefore, this paper recommends further study on the practical implications of applying these techniques to identify expectation gaps.

### 6.3.2 - Stakeholder salience

The evaluation of stakeholder salience is the second process of information gathering that the analysis examines. Not all reputational threats turn into a crisis and therefore, it is important to evaluate the salience of the stakeholder. Otherwise, the crisis practitioner does not know which threats to focus on. Coombs does not specifically provide any methods to determine the stakeholder salience, which makes it highly relevant to examine if big data analytics can assist the crisis practitioner. In this regard, the big data technique 'social influence detection' has potential to measure a stakeholder's power on social media. Although, the technique is limited in measuring power because online power is a small part of the power a stakeholder can hold over an organization.

In contrast to power, there is no way to determine legitimacy and willingness of a stakeholder with big data analytics. This would require an understanding of the psychic systems. The complexity of analyzing communication in order to understand the willingness or legitimacy of a stakeholder is far too complex for big data analytics. In this regard, machines are inferior in analysis meaning and communication, when it comes to small amount of data. As a result, big data analytics is neither able to determine the impact or likelihood score.

Statistically, it is more likely that a potential threat arises from a stakeholder with low power. Since it is difficult to foresee where a threat arises, it does not provide any value for the crisis practitioner to apply social influence detection before a threat is perceptible. Due to the limited time to take action when a threat arises, it is difficult to apply social influence detection. This technique takes time and resources to install. Furthermore, there is no guarantee that social influence detection is better at evaluation online power than human interpretation. Based on the analysis, this paper does not see much potential of applying social influence detection to determine the salience of the stakeholders, since it is very problematic and little prospects for reward.

### 6.3.3 - Situation awareness

When a potential crisis threatens, it is important for the crisis practitioner to create situation awareness in order to make better decisions. The information gathering process required to create situation awareness is part of the crisis recognition stage. It involves an understanding of the stakeholders' perceptions of the situation, a comprehension of the situation and the ability to project future states.

The use of sentiment and aspect analysis has the potential to give the crisis practitioner an overview of the perceptions of the stakeholders. The techniques are not able to provide a thorough description of the communication. However, they might assist the crisis practitioner in gaining an overview of discussed topics and the sentiment of these topics. Currently, there are no other methods for the crisis practitioner to obtain this information in Coombs' framework due to the short time-frame. The use of aspect and sentiment analysis cannot stand-alone. Nevertheless, this paper believes that the techniques have the potential to assist the crisis practitioner and recommends further study on the practical implications.

Coombs does not provide detailed information about what kind of information the crisis practitioner requires, when comprehending the situation. The information varies from crisis to crisis and is very case-specific. Descriptive analytics might be able to provide some valuable information, however the use of structural data cannot grasp concepts such as opinions and perceptions. Instead, it is useful, when analyzing trends and correlations. It is difficult for the analysis to evaluate whether big data analytics is useful for the crisis practitioner because it is very case-specific. Furthermore, Coombs is quite vague in his description of areas where descriptive analytics may be of use. There might be potential in using descriptive analytics in some crises, while not in others. However, it was not possible for the analysis to analyze the use of descriptive analytics in depth.

The projection of future states for social phenomena is problematic for big data analytics. Crises can quickly change radically and it is difficult for big data analytics to foresee these changes because they correlate with the stakeholders' perceptions of the situation. The only big data technique that involves projections of future states is predictive analytics, which analyze structural data. The use of structural data makes it very difficult to grasp the complexity of perceptions. In this regard, the value of applying big data analytics to project future states of crises is limited.

## 6.3.4 - Social media monitoring during a crisis

The analysis concludes that big data analytics can contribute to two out of three important information needs, when performing social media monitoring during a crisis. These two information needs are about what the dominant topics are and whether stakeholder comments are favorable or unfavorable. The information need big data analytics cannot contribute to is what stakeholders are actually saying during a crisis. No big data technique can cope with such a complex task. The big data techniques that contribute to this area are sentiment analysis, aspect analysis and real time analytics.

All three techniques have some contributions to social media monitoring during crises. Combined they can assist crisis practitioners with determining if sentiments are positive or negative towards the organization and what the dominant topics are. The three techniques can even determine the sentiments on a specific topic in real time. Compared to manual social media monitoring, big data analytics certainly have some contributions in terms of providing overview and fast analysis about some of these information needs. However, big data analytics cannot understand exactly what stakeholders are communicating. For this reason, manual social media monitoring still beats the technology on this specific information need.

## 6.3.5 - Crisis evaluation

The crisis evaluation is the final information gathering process, which the paper analyzed. At this stage, it is important to measure the impact of the crisis and evaluate the crisis practitioner's performance during the crisis. By the help of Coombs' framework, this paper identifies four areas of information gathering, namely: crisis records, stakeholder feedback, the word of mouth and media coverage.

### 6.3.5.1 - Evaluating the crisis team's performance

The use of crisis records involves internal data and therefore, this paper cannot determine the available data. It becomes very case-specific and it is difficult for this paper to generalize any analytical conclusions about the contributions of big data analytics, when analyzing crisis records.

It is difficult for big data analytics to analyze stakeholder feedback in depth to such an extent that it is valuable for the crisis practitioner. Big data analytics is not able to choose which stakeholders groups to focus on. Furthermore, it is not possible to direct the communication to topics of interest that are valuable at the post crisis stage. Big data analytics is limited to use available data. In this regard, there is little potential to use big data analytics to analyze stakeholder feedback.

The use of big data analytic to explore the word of mouth does not provide much value for the crisis practitioner either. At this stage, it is important to understand how the stakeholders perceive the organization in detail. Since there is little time pressure for the crisis practitioner, he has time to read the communication more thoroughly. Big data analytics is not able to analyze the communication to the extent that makes it relevant for a crisis practitioner. The techniques require more development to analyze the meaning of the communication.

This paper identifies a great potential for big data analytics to assist the crisis practitioner in tracking the media coverage. Within the field of big data analytics, analysts apply the big data technique 'information extraction' to detect quotations and spokespersons in multilingual media. In this regard, there is a high probability that the same methods can assist the crisis practitioner in measuring the media coverage of online media. Furthermore, the transcript-based approach uses audio and video data and transcribes it into text. By applying this technique, the crisis practitioner might be able to identify the media coverage of online audio and video as well. Coombs does not provide any methods to track the media coverage and therefore, these techniques have great potential to assist the crisis practitioner.

## 6.3.5.2 - Measuring the crisis impact

The evaluation of the crisis impact involves measuring the financial impact and implications on reputation. Coombs only mentions the topic of financial ratios briefly, which makes it difficult for this paper to analyze if big data techniques have any potential to assist the crisis practitioner. Based on Coombs' framework, it is not possible to determine if the calculation of financial ratios involves big data. Without the presence of big data, there is no need to apply big data techniques, as other techniques are superior and more tangible to use.

When analyzing the crisis' impact on the organization's reputation it is preferable to measure the impact on both traditional and digital media. Big data analytics is limited to analyzing digitalized media, because traditional media is not easily quantifiable. Analysts use sentiment analysis to calculate the online reputation of organization. The crisis practitioner might benefit by applying the same methods to the discipline of crisis communication. However, Coombs argue that traditional methods of calculating reputation are more accurate and they can reach a wider range of stakeholders. Since time pressure is not a big factor at the post-crisis stage, traditional methods apply more easily to measure reputation.

The use of big data analytics for calculating reputation post-crisis is limited, because full proven methods already exist. Nevertheless, both traditional methods and big data analytics can co-exist for each their purpose, as they do not analyze the same data. Big data analytics is able to reach a greater amount of stakeholders. Furthermore, digital media store the data online, which makes it possible for big data analytics to calculate the reputation at any point in time, in contrast to traditional methods. For organizations that does not calculate reputation regularly, this is an advantage.

## 6.4 - Evaluation of viable data

In the examination of viable big data techniques, only few have the potential to assist the crisis practitioner within the five areas in the framework. The following section discusses how viable it is to analyze different kind of data with big data analytics for the crisis practitioner. This discussion builds on the previous findings in the analysis. Throughout the analysis, text analytics techniques seem the most viable in regards to the crisis communication & management discipline. Sentiment and aspect analysis are able to identify the stakeholders' sentiment towards the organization. These techniques are the only techniques that has potential and capacity enough to help the crisis practitioner understand the stakeholder perceptions.

The only technique that the paper found viable in audio analytics is the transcript-based approach. This technique transforms and reduces audio data into text data after which it is possible to apply text analytics. In this transformation, the pronunciation is lost. However, at its current state, audio analytics cannot provide crisis practitioners with any valuable insights. The same conclusions apply for video analytics. Video analytics cannot determine the content of video material. In this regard, video analytics is not able to determine or interpret meaning, interpretations and perceptions.

Techniques within social media analytics involve the connectivity between different stakeholders. Concerning stakeholder salience, there is some use of social influence detection. However, this paper evaluates that the output from this technique is quite small. In general, it is not relevant to analyze the individual connections between stakeholders within the five selected areas of crisis communication. This paper did not find indications that it is useful to analyze consecutiveness on such a micro level within crisis management.

Big data techniques using structural data such as numbers and values provide little use when analyzing stakeholder perceptions. The areas where descriptive analytics might be useful for the crisis practitioner are very case specific. Furthermore, Coombs is very vague in his description of those areas. As a result, this paper was not able to go more in depth with descriptive analytics. Predictive and prescriptive analytics seeks to foresee future outcomes by analyzing the past. This is difficult when analyzing social phenomena that can change rather quickly. Perceptions are very complex and unpredictable. By analyzing crises as social phenomena, it is difficult to apply methods that usually operate within a more static image of the world. With a more positivistic definition of crises, this paper would likely find more use for big data techniques using structural data.

# Bibliography


Aggarwal, C.C., Zhai, C. (Eds.), 2012. Mining Text Data. Springer US, Boston, MA.

Ansgar Zerfass, Piet Verhoeven, Angeles Moreno, Ralph Tench, Dejan Verčič, 2016. EUROPEAN COMMUNICATION MONITOR 2016. EACD European Association of Communication Directors, Brussels, www.eacd-online.eu.

Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J., 2011. Everyone's an influencer: quantifying influence on twitter. ACM Press, p. 65. doi:10.1145/1935826.1935845

Big Data Analysis Techniques | Slice and Dice [WWW Document], n.d. . Search Technol. URL http://www.searchtechnologies.com/big-data-analysis-techniques (accessed 5.13.16).

Cambria, E., White, B., 2014. Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]. IEEE Comput. Intell. Mag. 9, 48–57. doi:10.1109/MCI.2014.2307227

Chen, H., Chiang, R.H.L., Storey, V.C., 2012. BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. Mis Q. 36, 1165–1188.

Chen, W., Yuan, Y., Zhang, L., 2010. Scalable Influence Maximization in Social Networks under the Linear Threshold Model. IEEE, pp. 88–97. doi:10.1109/ICDM.2010.118

Colleoni, E., Arvidsson, A., Hansen, L.K., Marchesini, A., 2013. Measuring Corporate Reputation using Sentiment Analysis. Presented at the The 15th International Conference on Corporate Reputation: Navigating the Reputation Economy, New Orleans, USA, pp. 1–25.

Colleoni, E., Rozza, A., Arvidsson, A., 2014. Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data: Political Homophily on Twitter. J. Commun. 64, 317–332. doi:10.1111/jcom.12084

Coombs, W.T., 2015. Ongoing crisis communication: planning, managing, and responding, Fourth edition. ed. SAGE, Thousand Oaks, California.

Crisis Communications: What to Do When a Crisis Occurs [WWW Document], 2016. . reportbrain. URL http://www.reportbrain.com/about/crisis-communications-crisis-occurs/ (accessed 9.15.16).

Datafloq - Is Big Data The Key to Crisis Management? [WWW Document], n.d. . Datafloq. URL https://datafloq.com/read/big-data-key-crisis-management/9 (accessed 9.15.16).

Digital Reasoning, 2014. Unstructured data: A big deal in big data. Digital Reasoning.

Eccles, R.G., Newquist, S.C., Schatz, R., 2007. Reputation and Its Risks [WWW Document]. Harv. Bus. Rev. URL https://hbr.org/2007/02/reputation-and-its-risks (accessed 9.15.16).

Edelman, M., 1993. Contestable categories and public opinion. Polit. Commun. 10, 231–242. doi:10.1080/10584609.1993.9962981

Entman, R.M., 1993. Framing: Toward Clarification of a Fractured Paradigm. J. Commun. 43, 51–58. doi:10.1111/j.1460-2466.1993.tb01304.x

Ernst & Young, 2014. Big data - Changing the way businesses compete and operate. Ernst & Young.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. The KDD process for extracting useful knowledge from volumes of data. Commun. ACM 39, 27–34. doi:10.1145/240455.240464

Fisher, D., DeLine, R., Czerwinski, M., Drucker, S., 2012. Interactions with big data analytics. interactions 19, 50. doi:10.1145/2168931.2168943

Framing Theory, 2011. . Mass Commun. Theory.

Gandomi, A., Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. Int. J. Inf. Manag. 35, 137–144.

Gartner In., 2016. IT Glossary.

Google Scholar Citations [WWW Document], n.d. URL https://scholar.google.com/citations?view_op=search_authors&hl=da&mauthors=label:crisis_communication&before_author=Xdb3_7ICAAAJ&astart=0 (accessed 9.15.16a).

Hahn, U., Mani, I., 2000. The challenges of automatic summarization. Computer 33, 29–36. doi:10.1109/2.881692

Henderson, J.V., Storeygard, A., Weil, D.N., 2012. Measuring Economic Growth from Outer Space. Am. Econ. Rev. 102, 994–1028. doi:10.1257/aer.102.2.994

Hornstrup, C., 2005. Systemisk ledelse: den refleksive praktiker. Dansk psykologisk Forlag ; Eksp. DBK], Virum; Køge.

Hviid Jacobsen, M., Lippert-Rasmussen, K., Nedergaard, P., 2011. Videnskabsteori i statskundskab, sociologi og forvaltning. Hans Reitzel, Kbh.

Intel IT Center, 2013. Predictive Analytics 101: Next-Generation Big Data Intelligence. Intel IT Center.

Jagadish, H.V., 2015. Big Data and Science: Myths and Reality. Big Data Res. 2, 49–52. doi:10.1016/j.bdr.2015.01.005

Kneer, G., Nassehi, A., 1997a. Niklas Luhmann: introduktion til teorien om sociale systemer. Hans Reitzel, Kbh.

Krishnan, S., Raina, S., Aher, N., 2014. Real Time Audio-Based Search in Media Files Using Machine Learning. IJCAT J. 1, 507–511.

Laney, D., 2001. 3D Data Management: Controlling Data Volume, Velocity, and Variety. META Group, Technical report.

Lassen, N.B., Madsen, R., Vatrapu, R., 2014. Predicting iPhone Sales from iPhone Tweets. IEEE, pp. 81–90. doi:10.1109/EDOC.2014.20

Liu, B., 2012. Sentiment analysis and opinion mining., Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Luhmann, N., 1995. Social systems, Writing science. Stanford University Press, Stanford, Calif.

Neuman, W.L., 2000. Social research methods: qualitative and quantitative approaches, 4th ed. ed. Allyn and Bacon, Boston.

Neustein, A. (Ed.), 2010. Advances in Speech Recognition. Springer US, Boston, MA.

Parthasarathy, S., Ruan, Y., Satuluri, V., 2011. Community Discovery in Social Networks: Applications, Methods and Emerging Trends, in: Aggarwal, C.C. (Ed.), Social Network Data Analytics. Springer US, Boston, MA, pp. 79–113.

Ph.d, T.B.M., Kforum, R., n.d. 2710 k-chefer er fortabte [WWW Document]. URL http://www.kommunikationsforum.dk/artikler/Den-store-maaling-af-K-faget-i-Europa (accessed 9.15.16).

Pouliquen, B., Steinberger, R., Best, C., 2007. Automatic Detection of Quotations in Multilingual News. Presented at the PROCEEDINGS OF THE INTERNATIONAL CONFERENCE RECENT ADVANCES IN NATURAL LANGUAGE PROCESSING, pp. 487–492.

Regazzoni, C., Cavallaro, A., Wu, Y., Konrad, J., Hampapur, A., 2010. Video Analytics for Surveillance: Theory and Practice [From the Guest Editors. IEEE Signal Process. Mag. 27, 16–17. doi:10.1109/MSP.2010.937451

Ridley, D., 2008. The literature review: a step-by-step guide for students, Sage study skills. SAGE, London ; Thousand Oaks, Calif.

Rienecker, L., Stray Jørgensen, P. (Eds.), 2005. Den gode opgave: håndbog i opgaveskrivning på videregående uddannelser, 3. udgave. ed. Samfundslitteratur, Frederiksberg.

Shan, C., Porikli, F., Xiang, T., Gong, S. (Eds.), 2012. Video Analytics for Business Intelligence, Studies in Computational Intelligence. Springer Berlin Heidelberg, Berlin, Heidelberg.

Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C., 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. Standford Press Release.

Song, D., Schilder, F., Smiley, C., Brew, C., 2015. Natural Language Question Answering and Analytics for Diverse and Interlinked Datasets. 2015 Conf. North Am. Chapter Assoc. Comput. 101–105.

Tamhane, D.S., Sayyad, S.N., 2015. BIG DATA ANALYSIS USING HACE THEOREM. Int. J. Adv. Res. Comput. Eng. Technol. IJARCET 4, 18–23.

Thyssen, O., 2012. Det filosofiske blik: europæiske mestertænkere. Information, Kbh.

Timothy Coombs - Google Scholar Citations [WWW Document], n.d. URL https://scholar.google.com/citations?user=GaNF1oYAAAAJ&hl=da (accessed 9.15.16).

Tomaszewski, B.M., Robinson, A.C., Weaver, C., Stryker, M., MacEachren, A.M., 2007. Geovisual Analytics and Crisis Management. Presented at the Conference Delft, In Proc. 4th International Information Systems for Crisis Response and Management (ISCRAM), Netherlands, pp. 1–8.

Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. Science 211, 453–458. doi:10.1126/science.7455683

Vilas, K.S., 2013. Big Data Mining. Int. J. Comput. Sci. Manag. Res. 12–17.

YouTube Statistics [WWW Document], 2016. URL https://www.youtube.com/yt/press/statistics.html