DESIGNING A METHODOLOGY FOR MARKETING INTELLIGENCE SYSTEMS – THE CASE OF BRAND IMAGE DIAGNOSTICS

Research paper

Egger, Marc, Universität zu Köln, Cologne, Germany, egger@wim.uni-koeln.de

Volkmann, Gloria, Universität zu Köln, Cologne, Germany, volkmann@wim.uni-koeln.de

Schoder, Detlef, Universität zu Köln, Cologne, Germany, schoder@wim.uni-koeln.de

Abstract

In situations of information overload and complexity, consumers consult their existing knowledge regarding brands as a guide in consumption decisions. This knowledge manifests as brand association networks (BANs) in consumers' minds and reflects what the consumer thinks of when being confronted with a brand stimulus. BANs therefore characterize a brand's image that determines consumers' attitudes and behaviour. BANs serve as diagnostic instruments to explain a brand's success or failure and to plan or control marketing activities. Traditionally, BANs are elicited directly from consumers utilizing survey-based instruments. However, in a dynamic and interactive environment, user-generated content (UGC) is increasingly relevant for a brand's image and thus should be exploited for the elicitation of BANs. However, established elicitation instruments either follow another elicitation paradigm (i.e. surveys or interviews), or are unable to cope with volume, velocity, and variety of UGC as a big data source (e.g. content analysis). Hence, exploiting UGC for BAN elicitation requires the development of new, computer-supported instruments. Following a design science research approach, we contribute a novel methodology as our artefact to extract BANs from UGC using text-mining and network analysis. We evaluate our solution and demonstrate its utility for brand management on a study of automotive brands.

Keywords: brand-image, association-networks, UGC, User-generate-Content, Text-Mining.

1 Introduction, Problem, and Motivation

In times of information overload, and converging product and service attributes, consumers have difficulties differentiating between products and services (Esch et al. 2008; Keller 1993). In those situations of uncertainty or during the preparation of consumption decisions, consumers consult their existing knowledge regarding brands (Aaker 1991; Collins and Loftus 1975; Esch et al. 2005; Keller 1993). This brand knowledge determines customer-based brand equity when consumers react more or less favourably to elements of the marketing mix of a brand in relation to an unbranded product or service (Keller 1993). Customer-based brand equity occurs when consumers have strong, favourable, and unique brand associations in their mind (Keller 1993; Keller and Lehmann 2006) and provides diagnostic information on the emergence of finance-oriented brand equity (Esch 2014) (represented as consumers' accepted price difference between a branded and an unbranded but equal product or service (Esch 2013)). To create competitive advantage and to increase finance-oriented brand equity, organizations have to develop strategies to control customer-based brand equity (Esch et al. 2008; Keller 1993). Before those strategies can be developed, a conceptualization and operationalization of measurement items is necessary. According to Keller's conceptualization (Keller 1993), customer-based brand equity is determined by brand knowledge in consumers' mind. Brand knowledge can be conceptualized being composed of brand awareness (e.g. consumers' performance to recognize and remember the brand in different situations) and brand image (Keller 1993) which manifests in networks of brand associations in consumers' mind. This perspective of brand image based on association networks is theoretically grounded in the human associative memory model (Anderson and Bower 1980) and the spreading activation theory (Collins and Loftus 1975) from cognitive psychology. This cognitivist perspective describes human semantic memory as a network of interconnected nodes where nodes store information, and edges of different strengths describe activation probability patterns. The activation of an informational node, e.g. originating from a brand stimulus, spreads through the network and activates other nodes according to the strength of the connecting edges. The activation pattern finally defines which information is retrieved from memory and comes to consumers' mind. This associative memory perspective has been widely accepted and adopted within marketing research (Esch 2014; Keller 1993; Keller and Lehmann 2006; Krishnan 1996). Here informational nodes in consumers' memory are represented by brand associations emerging from direct (e.g. usage) and indirect experiences (e.g. Word-of-Mouth (WoM)) of consumers with a company's offerings (Esch et al. 2008; Keller 1993; Krishnan 1996) and represent product- and non-product related associations, attributes, benefits, and attitudes (Keller 1993). Activation of those nodes describes which associations come to consumers' mind when they are confronted with a brand-related stimulus. Therefore, brand association networks serve as the foundation to understand consumer attitudes and preferences as the determiner of customer-based brand equity (Henderson et al. 2002; 1998; Keller 1993). Henderson et al. (1998; 2002) demonstrate the diagnostic power of brand association networks for branding mechanisms such as brand parity, confusion, and dilution. However, before brand association networks can be used for the derivation of strategies to increase customer-based brand equity, they have to be elicited. Various instruments have been proposed for which Farsky (2007) provides a comprehensive overview. Common to those instruments is that they elicit brand association networks directly from consumers by reactive survey- or interview-based research. However, today electronic word-of-mouth (eWOM) communication opens up new opportunities for brand association elicitation. eWOM can be defined as "any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet" (Henning-Thurau et al. 2003). A vast amount of the content that is produced by consumers during eWOM communication activities is text-based, voluntarily published, publicly visible on the Internet, and typically termed user-generated content (UGC). As UGC also covers consumers' attitudes, feelings, and opinions towards brands (Decker and Trusov 2010; Egger and Lang 2013; Netzer et al. 2012), UGC represents a novel and rich data source for the elicitation of brand association networks. Using UGC as a data source for brand association networks seems to be especially relevant as consumers increasingly consult other consumers' opinions from UGC (e.g. from product reviews) during

their purchase decisions (Adjei and Noble 2010; Gensler et al. 2015; King et al. 2014). Thus UGC does not only provide insights into the brand image of a brand's customers, but might also affect how other consumers perceive a brand (Gensler et al. 2015) which also fosters a brand's market performance. The relevance of UGC for the performance of brands in the marketplace has also been shown quantitatively. Dhar and Chang (2009) observe a correlation between sales numbers of newly published music albums and the number of blogs mentioning those albums. Likewise, Xiong and Bharadwaj (2014) show that UGC on blogs, forums and product review communities better predicts sales figures than traditional media or search engine volume. These contributions quantitatively show that UGC is relevant for brands and has impact on real-world consumer behaviour that influences a brand's performance. Nevertheless, a qualitative explanation on the reasons of consumers' behaviour and market performance remains open. Therefore, brand association networks should be elicited from UGC, as those can serve as a qualitative diagnostic tool on brand image explaining the reasons for real-world consumer behaviour, such as consumption decisions and finally explain the emergence of customer-based brand equity. Thus, using UGC as a data source for the elicitation of brand association networks for brand image diagnostics is relevant and might complement traditional research approaches, especially in a dynamic and interactive environment (Gensler et al. 2015). However, as brand associations in text-based UGC are represented by words and phrases, those have to be identified and extracted. Applying traditional approaches such as content analysis or open coding to UGC would require the researcher to read every text and manually identify brand associations and interconnections. Unfortunately, the volume and velocity of UGC creation as well as the scattered occurrences across the Web would lead to enormous efforts for traditional content analysis approaches if not restricting analysis to a small sub-sample of the total UGC available. Additionally, due to increasing competition, cost pressure and shortened (product) life cycles, organizations are increasingly interested in new fast and cost-efficient research approaches (Urban and Hauser 2004). UGC bears potential for both. As UGC is publicly visible and readily available, no participants have to be recruited and/or interviewed, which saves on recruiting costs. UGC is stored electronically and can be accessed automatically, which saves elicitation effort while simultaneously increasing elicitation speed. UGC is available digitally and allows automation, e.g. by utilizing automated text-mining techniques which would decrease processing costs while increasing processing performance. Additionally, using UGC for the elicitation of brand association networks would also open up new diagnostic capabilities, which allows new substantive insights into brand image: The availability of historic archives of UGC on the Web would allow ex-post analyses, the velocity of UGC creation would allow (near) real-time snapshots of brand association networks, and the variety of meta-data (such as date, author, host) would allow ad-hoc segmentations of brand association networks to provide insights into a brand image from different perspectives. As a result, UGC can be considered as a new, relevant, and cost-efficient data source to elicit brand association networks for brand image diagnostics that offers additional analytic capabilities but presently cannot be exploited sufficiently using traditional research instruments such as content analysis. Therefore, exploiting UGC as a data source for the elicitation of brand association networks leads to two research opportunities located at the intersection of marketing and information systems research. First, the elicitation of brand association networks from UGC requires novel (semi)automated research methods as traditional approaches are unable to catch up with the volume, velocity, and variety of UGC as a data source. Second, in a dynamic and interactive environment (Gensler et al. 2015), brand managers need information systems allowing them to assess attributes of their own and their competitors brand image ad-hoc and in short timeframes for rapid strategy development leading to competitive advantages. The article at hand addresses those research opportunities.

While building on knowledge from marketing research, we develop a novel methodology that uses text mining and network analysis to elicit brand association networks from UGC. Next, we describe an artefact in the form of an information system that implements the methodology and helps (brand) managers assess their and their competitors' brand association networks in a short period and offers additional analytic capabilities. Thus, we contribute to the intersection of marketing and information system research by proposing a novel methodology that exploits a new data source for brand association networks by following a design-science research approach as we design, construct, validate, and

communicate an artefact, which addresses an important and presently unmet practice problem in organizational brand management. For the construction of the methodology which represents our artefact we follow the design-science research process proposed by Peffers et al. (2007) which organizes the design-science research process into activities within six stages These stages cover the following activities: 1) identify the problem and motivate, 2) define the objectives of a solution, 3) design and development, 4) demonstration, 5) evaluation, and 6) communication. Therefore we structure our design-science research project according to the process proposed by Peffers et al. (2007). While this first section identified and motivated the problem of exploiting UGC by designing an information system serving as a diagnostic tool to brand image, Section 2 identifies objectives from related literature and transfers them to design objectives. Section 3 then describes our solution to elicit brand image from UGC, while Section 4 conducts a comprehensive study on automotive brands to demonstrate and evaluate our artefacts' brand image diagnostic capabilities. Section 5 then provides a discussion and gives an outlook to further research.

2 Objectives of the Solution

We understand our proposed solution as a methodology which we instantiate to illustrate its utility. We derive the objectives for our solution from first pointing out differences to previous research. Previous research: 1) either follows another elicitation paradigm that cannot be applied to UGC (i.e. surveys or interviews), 2) only uses a sample of the total UGC available (e.g. specific forums, blogs, product reviews), or 3) extracts brand-relevant information but no brand association networks. Brand association networks have been elicited from survey and interview data (Brandt et al. 2011; John et al. 2006; Schnittka et al. 2012) but not from UGC. Networks of product and brand names were extracted from text-based UGC for market structure analysis (Netzer et al. 2012). Text mining was applied on forums, blogs, and tweets to elicit rankings of brand associations (Crawford Camiciottoli et al. 2014), product features and/or general attributes (Liu et al. 2005; Popescu and Etzioni 2005) but no network structures. However, to be consistent with the theoretical concept of brand image, the elicitation of brand associations' interconnections is especially relevant as those describe how an activation of an association is spread through the association network (Collins and Loftus 1975). Gensler et al. (2015) notice this gap and are the first who consistently consider the theoretical understanding of brand image for the development of an automated methodology to elicit brand association networks from UGC. Although Gensler et al. (2015) were able demonstrate the diagnostic capabilities of their approach for brand management, the applicability of their approach is limited. First, Gensler et al. (2015) do not analyse unstructured text (which the bulk of UGC consists of). Second Gensler et al. (2015) only focus on a very specific characteristic of one product review platform, which represents only a sample of the total available UGC. Hence we derive the following objectives for our solution. From research on the theoretical foundations of the brand image and brand equity construct (Aaker 1991; Esch 2014; Keller 1993; Krishnan 1996), we define the first objective of our solution as **theoretical consistency (I)**. This means that the solution has to operate in a way to be consistent with accepted theories on brand image to serve as a diagnostic tool. Therefore the solution has to aim for the elicitation of brand association networks that allow being characterized according to the type, strength, favourability, and uniqueness of interconnected brand associations (Brandt et al. 2011; Henderson et al. 1998; Keller 1993; 2003; Keller and Lehmann 2006; Krishnan 1996).

Furthermore, related contributions that utilized UGC (see above) all focused on specific instances of UGC (e.g. forums, blogs, or product reviews) from which we derive the second objective: **source agnostic (II)**. This means that our solution might be technically applied on any kind of text-based UGC. Thus, it mustn't utilize any specific characteristic of a certain data source (e.g. utilizing the semi-structured product summarizations on some product review communities like Decker and Trusov (2010) or Gensler et al. (2015)). When the solution is potentially able to operate on any kind of text-based UGC (source agnostic), the solution should attempt to take as much UGC as possible into consideration and not preselect certain data sources (e.g. certain forums, blogs, social networks) to prevent a selection or sampling bias. Therefore we define the third objective for the solution to be **content ho-**

listic (III). A solution that is content holistic and attempts to include as much UGC as possible is faced with a practical problem as data quality on the Web differs. So, a solution that uses data from all over the Web has to address the fourth objective of **noise prevention (IV)** due to varying data quality, fraud, and irrelevant data. From meeting the objectives as described above, we primarily aim for a methodological contribution at the intersection of marketing and information systems research: We develop a novel methodology to elicit brand association networks from UGC that – in contrast to previous research – both captures a complete understanding of the brand-image construct (theoretical consistency) and does not restrict its scope to a specific instance of UGC such as product reviews, forums, or blogs (source agnostic, content holistic) to prevent sampling biases. Furthermore, we contribute to the ongoing discussion in design-science research as we follow a design-science research process and finally instantiate an artefact representing an information system for branding strategy and controlling. We evaluate the utility of our artefact by demonstrating its ability to capture and clean text-based UGC from the Web (noise preventing) and derive attributes that help comprehensively characterizing a brand's image both for different consumer groups as well as over time (by strength, favourability, uniqueness, and interconnections of brand associations; i.e., theoretical consistency).

3 Design and Development

In the following, we describe the methodology of our artefact to elicit and construct brand association networks from UGC. The proposed solution follows an elicitation process similar to the process Gensler et al. (2015) propose. 1) The research subject has to be defined and data collected; 2) before data cleaning is performed; 3) afterwards, brand associations are extracted from UGC. For each of those brand associations, 4) the sentiment, 5) strength, and 6) interconnections are determined.

3.1 Definition of Research Subject and Data Collection

First, the researcher has to define the research subject and its scope. The research subject might be a general brand or product brand for which text-based content from the Internet shall be collected. We propose using search engines such as Google or Bing to retrieve references to documents that relate to the research subject as we explain in the following. First, search engines' indexes are large as their primary purpose is to provide a maximum coverage of online available content. Second, search engines retrieve results fast, so they are able to be applied ad-hoc and in short timeframes. Third, search engines are targeted to consumers who query for information, so search engines' retrieved results best represent information that potentially also influences consumers during their personal search processes. Common to search engines is their Boolean-based query language where the user defines a query using keyword combinations in a way that the query best characterizes documents the user intents to find. When integrating search engines into other information systems, either application programming interfaces (APIs) might be used if available, or Web scrapers have to be developed. Those utilize the search engines' public search interface, insert the intended query into the search form, automatically read out URLs from the search-engine-provided result set and store them to a common database. Within our described artefact both a Web-search API or a Web scraper can be "plugged in" as long as the Web scraper or API allows defining a Boolean search query. From the collected URLs, our artefact retrieves the HTML content (in the following: HTML documents) from the corresponding Web page and stores it to a database.

3.2 Data Cleaning

To address the objective of noise prevention, collected HTML documents have to be cleaned, as they might contain unwanted, duplicate, or non-user-generated content (non-UGC). We propose a three-stage cleaning process. In the first stage, irrelevant boilerplate elements (such as navigation, header, footer elements) have to be isolated from information carrying content. Every collected HTML document therefore might consist of multiple boilerplate elements as well as multiple text elements that might carry information. In our artefact, we apply the boilerplate detection approach proposed by

Kohlschütter et al. (2010). This approach uses general Web-document attributes such as the link-to-text ratio to differentiate boilerplate from textual content. This approach can be considered robust with regard to heterogeneous input documents, which is especially relevant when collecting HTML documents from all over the Web as our artefact does. In the second cleaning stage, duplicate content has to be cleaned. Duplicates occur when the same content is found on multiple documents – e.g. standard and accessible Web pages for screen readers, print views, or product descriptions. Absolute duplicates can be easily detected and cleaned using character-wise comparison, while cleaning near duplicates requires more sophisticated approaches such as the fingerprinting-based ones described in Theobald et al. (2008). In the third cleaning stage, non-UGC is cleaned to create a dataset that solely consists of documents that reflect consumers' perspectives. We follow Egger et al. (2015) who propose and formulate a binary text classification problem to be addressed by supervised machine learning (Egger et al. 2015). For training the machine learning classifier, our artefact allows the researcher to manually annotate a random subsample of a few hundred collected documents according to UGC vs. non-UGC. Based on this annotated sample, a support vector machine (SVM) based classifier (Cortes and Vapnik 1995) is trained following the suggestions from Egger et al. (2015). The classifier is then applied on the whole of collected documents to identify non-UGC, which is erased afterwards.

3.3 Extraction of Brand Associations

In UGC documents, brand associations manifest as words and phrases. To extract those words and phrases that potentially represent brand associations, we need to know how consumers verbally express their brand associations. To the best of our knowledge, no research previously has formalized linguistic characteristics of brand associations, which would help the construction of automated brand association elicitation systems. However, Lawson (1998) considers brand associations being typically represented by concepts that describe fundamental attributes, similar products and topics, product usage scenarios or summarizing evaluations. Furthermore Kleining (1959) states that brand image associations are best described using words one would also use to describe a person. Obviously, when describing a person one would basically use adjectives and verbs to describe characteristics of the person (e.g. nice, good looking, ugly). Nouns would be used to refer to certain attributes (e.g. hair, hands, face) while adjective-noun combinations represent information chunks to characterize the person. (e.g. "good swimmer", "nice hair", "ugly person", "large hands"). We consider this information for the extraction of brand associations, which consists of tokenization, part-of-speech (PoS), PoS-sequence extraction, and sentiment analysis. First, we use tokenization (Feldman and Sanger 2006) to segment each UGC document into smaller textual entities. Tokenization on punctuation marks is used to receive single sentences. Ambiguous sentence endings (e.g. abbreviations, digit separators) are detected and omitted using dictionaries and handcrafted rules following Grefenstette and Tapanainen (1994). Afterwards, each sentence is tokenized into single words using whitespace characters. As we cannot assure that consumers only refer to the research subject (i.e. brand) within their published content, we restrict analysis to those sentences where hints to the research subject are found (e.g. mentioning the brand). Every sentence is then mapped to a secondary representation consisting of a sequence of PoS-tags (e.g. noun, adjective, etc.). On this secondary representation, our artefact allows the researcher to define PoS sequences that are used to extract brand association candidates. The researcher thereby has to provide the PoS sequence to the artefact and also tell which PoS tags should compose the brand association. As an example, the researcher might define the PoS sequence of <article, noun, verb, adjective> and define that the noun and the adjective represent the brand association. As a result, the artefact would extract the brand association <delicious burger> from the sentence <the burger is delicious>. For each of the UGC documents, the extracted brand associations are stored to a database.

3.3.1 Favourability of Brand Associations

The emotional orientation (e.g. positive, negative, neutral) of brand associations is of strong interest. Keller (1993) states that customer-based brand equity does not only emerge from strong and unique but also from favourable (i.e. positive) brand associations in consumers' minds. Therefore, extracted

brand associations should be also characterized according to their sentiment. Liu (2014) provides a comprehensive overview of sentiment analysis approaches. With regard to our solution, so-called aspect-oriented sentiment analysis is relevant as we want to assign sentiments on the level of single words and phrases. As the research field of aspect-oriented sentiment analysis develops rapidly with regard to detection accuracy, our artefact can be easily adapted to include recent state-of-the-art aspect-oriented sentiment analysis techniques. For illustrative purposes, in our instantiation of the proposed artefact, basic dictionary-based sentiment analysis is implemented. Hereby a referencing sentiment dictionary like SenticNet (Cambria et al. 2014) is used to transfer the sentiment of an adjective found in the dictionary to the relating noun. As a result of this sentiment analysis stage, for each UGC document a set of brand association candidates (e.g. "leading technology", "threatened privacy") is received. The noun hereby represents the brand association (e.g. technology, privacy) while the referencing adjective (e.g. leading, threatened) is used as sentiment driving aspect.

3.3.2 Strength of Brand Associations

The strength of a brand association determines how fast and easy an association can be retrieved from the memory of a consumer (Keller 1993). Therefore, the strength of brand associations is an important attribute for the evaluation of brand association networks (Keller 1993). To compute the strength of an association, we construct a 2-mode network of associations and documents. For each of the collected UGC documents, a document node is created. For each extracted association furthermore an association node exists. If one association has been extracted from different documents, the corresponding association node is connected to each of the document nodes via a document-to-association edge (see Figure 1). The degree centrality of an association node describes the number of documents where the association is mentioned and indicates an associations' strength (Scott 2012). To allow comparisons across research subjects, degree centrality is normalized from the total number of documents.

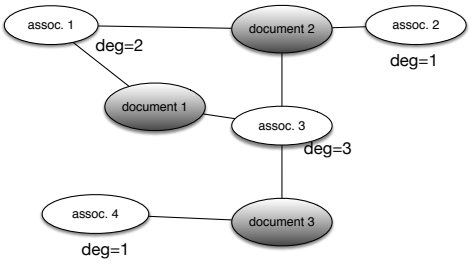

Figure 1. 2-mode Network to Determine Associations' Strength.

3.3.3 Uniqueness of Brand Associations

The knowledge about the uniqueness of brand associations allows brand managers discover differentiating attributes and to effectively position their brands in the marketplace (Gensler et al. 2015). Gensler et al. (2015) propose to define the uniqueness of brand associations by the difference of an associations' degree centrality. While this indicator treats every association as equally important, we additionally propose to also consider the strength of an association. The best differentiating associations are those which are mentioned frequently for one brand while being mentioned less frequently for the other brand. Therefore, associations that occur nearly equally frequently for two brands (no matter if frequent or less frequent) do not differentiate those brands. Term frequency/inverse document frequency (TF/IDF) (Ramos 2003) is an indicator initially proposed for information retrieval purposes that describes this relation of uniqueness. Within our solution we utilize TF/IDF using an associations' degree centrality for the term frequency to discover most differentiating associations – hence those associations that are most unique.

3.3.4 Connections Among Brand Associations

The knowledge of how associations are related in consumers' minds allows discovering how activation spreads across the association network (Keller 1993), which is especially important when designing marketing strategies. In consumer mapping approaches, consumers are directly asked to indicate their associations' interconnections (John et al. 2006). Unfortunately, this is not possible when passively analysing UGC. Therefore analytical mapping approaches seem promising, as those use network analysis to conclude relations between associations (Ahn 2013; Henderson et al. 2002). Therefore, similar to Teichert and Schöntag (2010), we assume that all associations consumers state in their utterances are somehow related to each other. Based on this assumption, our approach creates a complete graph of associations for each document. Using the entire document collection, co-occurring associations are counted and represented by a weighted edge. The weight of the edge corresponds to the number of equal co-occurring associations. Figure 2 illustrates the resulting network based on the complete graphs from different documents. As more frequent associations have a higher probability to co-occur with less frequent associations, a correction factor is applied to highlight those co-occurrences that have been observed more frequently than one would expect stochastically. Assuming statistical independence, the probability of two associations to co-occur stochastically amounts to the product of the relative occurrence frequency (i.e. normalized degree centrality) of those associations. This expected probability is subtracted from the normalized co-occurrence (quotient of absolute co-occurrence of both associations and the occurrence of the less frequent association).

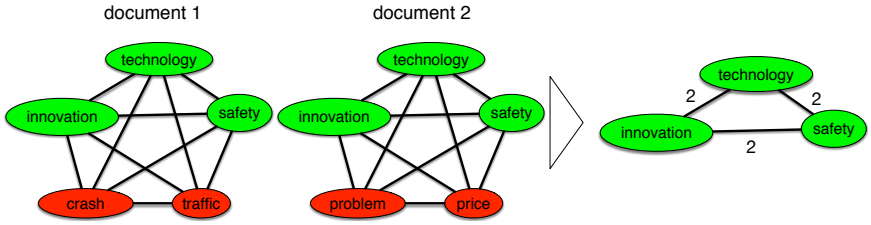

Figure 2. Individual and Complete Document and Consensus Networks.

3.4 Mapping the Brand Association Network

To construct the final brand association network, both intermediate networks from Section 3.3.2 and 3.3.4 are combined. While the network from Section 3.3.2 contributes each association's strength, the network from Section 3.3.4 contributes interconnections among associations. To visualize the association network we propose using concept maps (Novak and Gowin 1984). Concept maps were developed to communicate information, to construct and externalize knowledge to foster learning and understanding (Leake et al. 2004). Visualizing association networks as concept maps seems adequate, as brand association networks are intended to create a common understanding on a brand's image within organizational brand management and should also be easily understandable for senior management. Within concept maps, a concept represents a perceived regularity of objects or events which can be assigned with a label (which most of the time is represented by a single word) (Novak and Cañas 2008). The centre of the concept map represents the research subject (e.g. target brand). Directly connected to this centre, the concept nodes (that represent brand associations) are located. The distance of a concept node to the centre is inversely proportional to the strength of the brand association by means of normalized degree centrality (see section 3.3.2). The concept nodes label also maps the association strength in per mille while the colour describes favourability (red: negative, green: positive) as determined by sentiment analysis (see section 3.3.1). The lines between concept nodes describe the structure of the association network. To foster interactive exploration, the instantiation of our proposed artefact allows the user to define cut-off values on the scope of the brand association network. This scope can be defined according to the minimum normalized degree centrality of an association. Furthermore the user might define an N to discover the association network of the top-N associations by means of their strength (represented by the normalized degree centrality).

3.5 Segmenting the Brand Association Network

The methodology that represents our artefact maps UGC documents and associations as interconnected network nodes. For the instantiation of the artefact this means that for each association, all documents where the association occurs can be retrieved and vice versa. As a result, in information systems such as Business Intelligence Systems, ad-hoc network segmentations become possible. These would allow various applications as segmenting brand association networks by a certain association, allowing the creation of different perspectives on the network. As an example, it would be possible to separately investigate the brand association network of consumers who perceive "quality" as favourable and those that perceive "quality" as unfavourable. This would allow for discovering and comparing differences in consumer perception based on associations. Furthermore, segmentations could be applied on any meta-data available to the documents to create network partitions. For example, using the publication date would allow for discovering and comparing changes in brand association networks over time. When using the host-name (such as kitchen-forum.com) as the segmentation criterion, one would be able to investigate differences in perception among consumer groups such as comparing the association networks from consumers on a household forum with those on an automotive forum. When additional meta-data is available (such as gender, age, or location) demographic segmentations also become possible.

4 Demonstration, Evaluation, and Communication

In the following, we illustrate and demonstrate our proposed methodology as the design artefact and solution (see Peffers et al. 2007). We instantiated our proposed methodology as Java software which allows organizations to assess their brand association networks. To date, the software instantiation of our proposed methodology was applied in over 60 commercial analysis projects in 2017. By demonstrating this software instantiation, we show that our solution (artefact) meets the objectives from Section 2. Following a design-science research process (Peffers et al. 2007) we further illustrate the utility for brand-image diagnostics to assess brand image and to derive additional substantive insights. In the following we use the instantiation of our proposed methodology to conduct an illustrative study. We define the research subject as being automotive brands. We use the automotive brands Volkswagen, BMW, and Audi as our research subject because cars can be considered high-involvement products and the selected brands have been discussed controversially and emotionally in the past. While Volkswagen was faced with a major image crisis due to the manipulation of its diesel engines, Audi and BMW are close competitors targeting similar consumer groups. As a result we expect to find a lot of eWOM communication online. For each of our research subjects, we used our instantiation to collect documents from the Web. Thereby, we developed a Web crawler which inserts a defined search query in the Google search engine, visits and downloads HTML content from the retrieved URLs, cleans up boilerplate elements, and stores a text representation of the documents to a common database (see Section 3.1). As the search query, we used the brand names "Audi", "BMW" and "VW OR Volkswagen". As a result, our artefact collected 388,584 text-based documents for BMW (VW: 268,845; Audi: 206,331) published between January 2015 and May 2016. In total, our instantiation has collected text documents from 52,060 different websites (e.g. car-forum.com) for BMW (VW: 46,615; Audi: 33,977). This targets the artefacts' objective of being content holistic (maximizing the number of sources to be considered) and source agnostic (as our artefact was able to collect text data from various types of websites such as company sites, forums, blogs, newspapers, and social networks).

Next, we evaluate the "noise prevention" objective (see Section 2) with regard to preventing non-UGC from introducing noise to the analysis process. To support the researcher, our instantiation provides a user interface of a random sample of the total collected documents to the researcher. The researcher then manually annotates this sample with regard to the characteristic of UGC and non-UGC to create a gold set, which represents the documents' correct classification. With regard to our demonstrating study, we annotated 400 documents (UGC: 172; non-UGC: 228). This gold set is then used to train an SVM-classifier to automatically segment UGC from non-UGC. UGC classification performance is then validated in a 10-fold cross validation on a linear SVM kernel. Following 10-fold cross valida-

tion, the gold-set is segmented into 10 equally-sized segments, where in 10 iterations, 9 of the segments are used for training and 1 segment is used for validating the classifier. For validation the metrics of accuracy, precision and recall are used. As a result of 10-fold cross validation on the illustrating study, we receive an accuracy of 90.60%, a precision of 88.37% and a recall of 91.57% on UGC classification. Based on these results, we conclude that the artefact meets the "noise prevention" objective with respect to cleansing non-UGC. As a result of data cleansing (see Section 3.2) we finally receive a document collection of 198,530 UGC documents for BMW (VW: 135,473; Audi: 101,051).

4.1 Brand Association Networks for Brand Image Diagnostics

After having evaluated our artefact targeting the objectives "content holistic" and "noise preventing", we now evaluate "theoretical consistency" and show the utility of the instantiated artefact being used as a brand image diagnostic instrument. Characterizing a brand's image requires the elicitation of those brand associations and interconnections that can be described by their type, strength, favourability, and uniqueness (Aaker 1991; Esch 2014; Gensler et al. 2015; Henderson et al. 1998; Schnittka et al. 2012). To illustrate the artefact's capability to characterize brand image, we utilize the artefact's instantiation to extract the brand association network (Figure 4) for Audi from our collected and cleaned data. Thereby following our proposed methodology in Section 3 to use the pattern <adjective, noun> on the secondary PoS-sequence representation of each document to extract associations like <fast, car> from the primary representation. Similar to the association networks described in Henderson et al. (1998; 2002) and Gensler et al. (2015) our brand association networks allow intra- and inter-network analyses. While the first relates to the description of one brand, the latter aims to compare different brands or sub-networks. With regard to intra-network analysis, our instantiated artefact elicits networks that can be characterized by the type, strength, and favourability of brand associations. Furthermore, network analyses on the structure of the association network allow for deriving insights into how an activation might spread across the network. Our instantiated artefact presents the association network using concept maps (see Section 3.4). Furthermore it allows the researcher to define how elicited associations are represented as concept nodes within the network. For example, the researcher might discover the association network from a different perspective when defining that a concept should be created from the noun or adjective component from an extracted association (that both consists of an adjective relating to a noun). When using this feature, the instantiated artefact also allows the researcher to retrieve the other relating association component interactively from selecting it within its user interface.

As an example, the network in Figure 4 (left) allows the researcher to interactively show related adjectives as an explanation of an associations' favourability, while the network in Figure 4 (right) allows the researcher to interactively retrieve attributes to which the association relates. This also allows characterizing brand associations according to their type. While the attribute network on the left (Figure 4) predominantly characterizes product-related (e.g. "design", "quality", "interior", "engine") and non-product related attributes ("brand", "price", "model" or "feature"), the network on the right shows associations that describe attitudes (e.g. "great", "top", "good", "best", "perfect", "amazing") and benefits (e.g. "fast", "powerful", "easy", "sporty", "luxurious"). Associations can also be characterized by their strength. Figure 4 shows that, apart from the brand-mentioning association "Audi", "car" represents the strongest brand association. Consumers initially think of cars when presented with the stimulus "Audi", which is valid since Audi is most famous for cars. Furthermore "car" is favourable. Our instantiated artefact allows interactively revealing related association components that describe why "car" is perceived favourable. Audi's cars are perceived as "best", "super", "great", or "fast". Similar, the third strongest association "Audi" (unfavourable) can be explained as consumers mention Audi being "expensive". Other strong brand associations are "luxury", "look", "design", "performance" and "quality" which allows managers to easily communicate a characterization of the Audi brand: Audi's stands for great but expensive cars. Performance, look, design, and quality are other favourable attributes consumer associate with the brand. After having characterized brand associations by type, strength, and favourability, our artefact also elicits relationships between brand associations (see Section 3.3.4). According to Collins and Loftus (1975), an activation of an association (e.g. originating

from a stimulus) spreads across the network while activating other associations to finally describe what comes to consumers' mind. This information is especially important for brand managers when planning marketing instruments by choosing stimuli that activate the maximum number of other positive associations. With regard to our illustration, we find interesting relationships between *design*, *performance,* and *technology*. One interpretation would be that consumers being reminded on Audis "design" also think of "technology" and "performance", which helps to *construct effective marketing campaigns*. Next to this qualitative assessment of network structures, network analysis might also discover interesting associations and structures. For example, betweenness centrality – defined by the number of shortest paths including the association (Scott 2012) – might reveal gatekeeping associations that connect sub-networks and clusters. Activating associations with high betweenness centrality would lead to the most diverse other associations being activated, too (Gensler et al. 2015; Henderson et al. 1998; 2002). Similarly, activating associations with a high closeness centrality – sum of shortest paths to all other associations (Scott 2012) – would cause activation to spread faster through the network, as associations with high closeness centrality have short paths to every other association. To foster these kinds of additional network analyses, the instantiation of our artefact allows the researcher to export association networks to the Pajek network analysis software (Batagelj and Mrvar 1998).

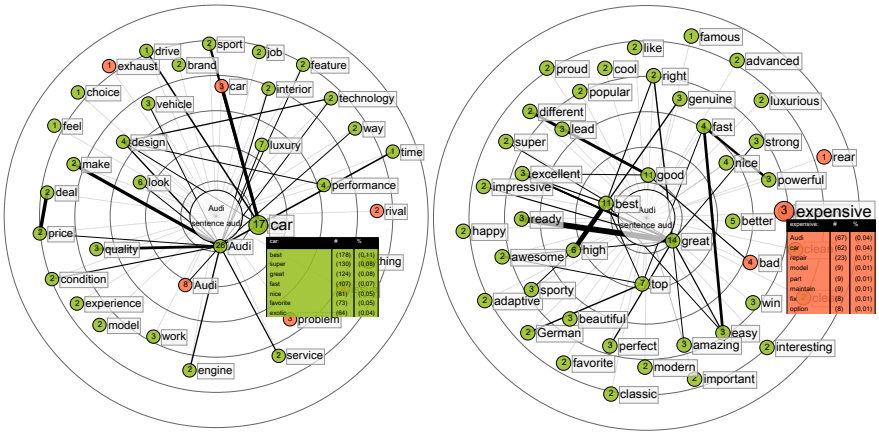

Figure 3. Attribute and Attitude Maps for Audi.

After having shown the ability of our artefact to characterize brand image from type, favourability, strength, and network structure, we now demonstrate how to characterize associations by their uniqueness. In doing so, we compute the TF/IDF-based uniqueness metric we proposed in Section 3.3.3 on the collected data for Audi and BMW. As a result, we find product-related associations to differentiate between both brands the most. The most differentiating associations for Audi are "Quattro" (Audi's name for four-wheel driving), R8 (a premium car model), VW (Audi is a member of the VW group, which consumers seem to notice), and other car model associations (such as A3, Avant, E-tron, TT, Sportback, and A6). BMW's most differentiating associations are: M (the premium product line), i3 (electric car), M3 (premium version of the three-series model), xDrive (BMW's name for four-wheel driving), and motorcycle (consumers seem to note that BMW also produces motorcycles, which Audi doesn't). Interestingly, we find only few associations that differentiate between both brands and do not describe car models (or sub-brands). While for Audi, "luxury" and "interior" describes the more unique associations, for BMW "driving (favourable)" is more unique. Thus, BMW's claim "sheer driving pleasure" seems to be noted from consumers. Nevertheless, BMW and Audi seem to be close competitors serving the market with similar offerings.

4.2 Additional Insights from Segmenting Brand Association Networks

After having illustrated the ability to characterize a brand's image, we now demonstrate additional brand-image diagnostics by the instantiations' interactive segmentation capabilities that emerge from the artefacts specific design. As an example, we illustrate how changes in association networks can be

traced over time by segmenting data by its publication date. This allows describing long-term effects in consumer perceptions or investigate how internal (such as marketing programs) or external events (product harm crisis) impact a brand's image. As an example, we conduct a study on the VW brand, which recently faced a major crisis threatening the VW brand image. In September 2015, the United States Environmental Protection Agency (EPA) found out that VW had manipulated its diesel engines' control software in a way that during laboratory emission testing, the engines emitted less nitrogen oxides than in real-life situations. When this incident became public, VW's CEO had to resign and VW had refit up to 11 million vehicles. We use our collected UGC (see Section 3.2) that includes date information and create two segmentations of VW's brand association network. The first segmentation is created from UGC published before September 2015 (54,203 documents), the other segmentation covers UGC published after September 2016 (61,998 documents). Table 1 shows the top 20 associations for both segments. It can be seen that the emission crisis is strongly reflected by the associations. The amount of unfavourable associations has increased from 6 to 8. New, unfavourable associations such as "scandal", "emission", "software", and "cheating" have emerged, while favourable associations from the before-crisis period disappeared from the top 20 (e.g. "design", "performance", "fun", or "experience"). The unfavourable association "Volkswagen" now ranks higher in the after period than the favourable one in the before period. Nevertheless, favourable attributes still remain (e.g. "car", "Volkswagen", "look", "vehicle", or "fine"), while others have not changed in order (e.g. "problem"). Finally, our results show that the VW emission scandal is strongly reflected in VW's association network, while favourable core associations still remain.

15/01-15/08 (before)	**type**	**degree**	**15/10-18/03 (after)**	**type**	**degree**
VW	PRO	0.0199	car	PRO	0.0122
car	PRO	0.0133	scandal	CONTRA	0.0103
Volkswagen	PRO	0.0069	Volkswagen	CONTRA	0.0094
problem	CONTRA	0.0041	Volkswagen	PRO	0.0084
car	CONTRA	0.0035	problem	CONTRA	0.0047
issue	CONTRA	0.0034	car	CONTRA	0.0043
Volkswagen	CONTRA	0.0025	emission	CONTRA	0.0043
vehicle	PRO	0.0024	issue	CONTRA	0.0040
thing	PRO	0.0023	look	PRO	0.0036
quality	PRO	0.0019	diesel	PRO	0.0035
engine	PRO	0.0018	vehicle	PRO	0.0029
design	PRO	0.0018	chance	PRO	0.0029
bug	CONTRA	0.0017	software	CONTRA	0.0026
performance	PRO	0.0017	fine	PRO	0.0026
part	PRO	0.0017	cheating	CONTRA	0.0023
fun	PRO	0.0016	thing	PRO	0.0023
experience	PRO	0.0015	way	PRO	0.0022
way	PRO	0.0015	engine	PRO	0.0021
job	PRO	0.0015	quality	PRO	0.0020
condition	PRO	0.0014	time	PRO	0.0019

Table 1. Top 20 Associations Before and After Emission Scandal

5 Discussion, Limitations, and Future Research

Our research objective was to develop a holistic and instantiable methodology as our design artefact that utilizes UGC to elicit brand association networks to assess a brand's image. The gap we intended

to fill emerges from previously proposed instruments which have either: 1) followed another elicitation paradigm which cannot be applied to UGC (e.g. surveys or interviews), 2) can't cope with the amount of UGC (e.g. content analysis), 3) considered only single-selected characteristics of brand image, or 4) restricted its scope to few selected UGC sources (e.g. product review communities or forums).

To address this gap, we followed Peffers et al. (2007) design-science research approach. We first defined the objectives for our solution (i.e. methodology), which has to be consistent with brand-image theory (theoretical consistency), mustn't be fitted to a specific instance of UGC such as product reviews (source agnostic), has to maximize the considered content (content holistic), and has to verifiably ensure data quality (noise prevention). We then proposed a methodology as our design artefact, which utilizes information retrieval, text mining, and network analysis to elicit brand association networks from UGC. We demonstrated the utility and diagnostic capabilities by instantiating our methodology and conducting a study on automotive brands, while evaluating against the defined objectives. Furthermore, we demonstrated the abilities of the instantiated design artefact to act as an ad-hoc brand image diagnostics system by its interactive segmentation capabilities performing an ex-post longitudinal study on the VW emission scandal.

The design artefact as our solution is (cost) efficient due to automation and allows processing large amounts of data in a short time, and is reliable as it can be applied independently by multiple researchers while leading to the same results. The results of the methodology are quantifiable and comparable as they rely on open and established metrics from network analysis. Furthermore our artefact fosters exploratory research due to its design that allows instantiations that offer interactive segmentation capabilities on any attribute that is available to the collected UGC. General limitations exist, as the human mind also stores other types of brand associations (such as sensory or visual brand associations), which are difficult to explicate (Supphellen 2000). Those will not be elicited by our methodology (i.e. artefact) as it considers only verbal associations from text-based UGC.

Furthermore the applicability of our solution depends on the availability of public UGC. Therefore it can only be applied to brands that consumers discuss online, which excludes smaller brands or brands targeted to business customers. As consumers on the Web often do not stick to one topic, our solution assumes that sentences where the brand is mentioned and relates to the brand, and therefore only considers those sentences. However, this assumption has to be questioned as in some sentences the brand is mentioned in passing, while other sentences do not mention the brand but refer to it (e.g. when using anaphora). As a result, our solution presently might exclude relevant sentences while also including irrelevant ones, which weakens the noise prevention objective. However, our approach represents a first approximation on intra-document brand relevance. Additional research should investigate how brand relevance can be determined on a (sub)sentence level. Nevertheless, associations elicited by our solution seem face-valid, but additional validation is necessary. Future research might address concurrent validity (Association 1954) by comparing the results of the proposed solution to those of an established brand association elicitation instrument, such as free elicitation (Olson and Muderrisoglu 1979).

Finally, despite the limitations, our solution represents the first comprehensive approach that exploits UGC for brand-image diagnostics which both aligns with branding theory, is holistic as it doesn't restrict investigation to a certain sample of UGC, and maps the entire elicitation process from data collection to visualization. Finally, we hope that our research encourages others to extend or further investigate our methodology from other research perspectives as well. While we hope that computer science researchers might contribute to refining technical mechanisms (e.g. intra-document brand relevance), behaviourist researchers might conduct explanatory research and investigate the relationship between the content and structure of brand association networks and market-oriented performance metrics, such as sales figures or finance-oriented brand equity measures.

References

Aaker, D. A. 1991. "Managing Brand Equity: Capitalizing on the Value of a Brand Name," New York.

Adjei, M., and Noble, C. 2010. "The influence of C2C communications in online brand communities on customer purchase behavior," Journal of the Academy of Marketing Science (38:5), pp. 634–653.

Ahn, H. J. 2013. "MINING TEXTS TO UNDERSTAND CUSTOMERS' IMAGE OF BRANDS," International Journal of Electronic Commerce Studies (4:1), pp. 141–144 (doi: 10.7903/ijecs.1122).

Anderson, J. R., and Bower, G. H. 1980. Human Associative Memory, Lawrence Erlbaum.

Association, A. P. 1954. Technical recommendations for psychological tests and diagnostic techniques (Vol. 51), Part 2 Psychological Bulletin.

Batagelj, V., and Mrvar, A. 1998. "Pajek-program for large network analysis," Connections (21:2), pp. 47–57.

Brandt, C., de Mortanges, C. P., Bluemelhuber, C., and van Riel, A. C. 2011. "Associative networks: A new approach to market segmentation," International Journal of Market Research (53:2), pp. 187–207.

Cambria, E., Olsher, D., and Rajagopal, D. 2014. "SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis,," Presented at the Twenty-eighth AAAI conference on artificial intelligence.

Collins, A. M., and Loftus, E. F. 1975. "A spreading-activation theory of semantic processing.," Psychological review (82:6), American Psychological Association, p. 407.

Cortes, C., and Vapnik, V. 1995. "Support-vector networks," Mach. Learn. (20:3), Springer, pp. 273–297.

Crawford Camiciottoli, B., Ranfagni, S., and Guercini, S. 2014. "Exploring brand associations: an innovative methodological approach," European Journal of Marketing (48:5/6), pp. 1092–1112 (doi: 10.1108/EJM-12-2011-0770).

Decker, R., and Trusov, M. 2010. "Estimating aggregate consumer preferences from online product reviews," International Journal of Research in Marketing (27:4), pp. 293–307 (doi: 10.1016/j.ijresmar.2010.09.001).

Dhar, V., and Chang, E. A. 2009. "Does chatter matter? The impact of user-generated content on music sales," Journal of Interactive Marketing (23:4), Elsevier, pp. 300–307.

Egger, M., and Lang, A. 2013. "A Brief Tutorial on How to Extract Information from User-Generated Content (UGC)," KI-Künstliche Intelligenz (27:1), Springer, pp. 53–60.

Egger, M., Lang, A., and Schoder, D. 2015. "Who Are We Listening to? Detecting User-generated Content (UGC) on the Web,," Presented at the ECIS 2015 Completed Research Papers, Münster, August 26 (doi: 10.18151/7217308).

Esch, F.-R. 2013. Moderne Markenführung: Grundlagen-Innovative Ansätze-Praktische Umsetzungen, Springer-Verlag.

Esch, F.-R. 2014. Strategie und Technik der Markenführung, Vahlen.

Esch, F.-R., Herrmann, A., and Sattler, H. 2008. "Marketing-Eine managementorientierte Einführung," Auflage, Vahlen Vl.

Esch, F.-R., Langner, T., and Rempel, J. E. 2005. "Ansätze zur Erfassung und Entwicklung der Markenidentität," in link.springer.com, Wiesbaden: Gabler Verlag, pp. 103–129 (doi: 10.1007/978-3-8349-4541-9_4).

Farsky, M. 2007. "Methoden zur Messung des Markenimages: State of the Art," No. 38, (Research Papers on Marketing and Retailing.), Hamburg: Research Papers on Marketing and Retailing.

Feldman, R., and Sanger, J. 2006. The text mining handbook: advanced approaches in analyzing unstructured data, Cambridge University Press.

Gensler, S., Völckner, F., Egger, M., Fischbach, K., and Schoder, D. 2015. "Listen to Your Customers: Insights into Brand Image Using Online Consumer-Generated Product Reviews," International Journal of Electronic Commerce (20:1), Routledge, pp. 112–141 (doi: 10.1080/10864415.2016.1061792).

Grefenstette, G., and Tapanainen, P. 1994. "What is a word, What is a sentence? Problems of Tokenization,," Presented at the Proceedings of COMPLEX, Budapest, Hungary: Citeseer, pp. 7–10.

Henderson, G. R., Iacobucci, D., and Calder, B. J. 2002. "Using network analysis to understand brands," Advances in Consumer Research (29), pp. 397–405.

Henderson, G., Iacobucci, D., and Calder, B. J. 1998. "Brand diagnostics: Mapping branding effects using consumer associative networks," European Journal of Operational Research 111, pp. 306–327.

Hennig-Thurau, T., and Walsh G. 2003. "Electronic Word-of- Mouth: Motives for and Consequences of Reading Customer Articulations on the Internet," International Journal of Electronic Commerce, 8, pp. 51–74.

John, D., Loken, B., Kim, K., Monga, AB. 2006. "Brand concept maps: a methodology for identifying brand association networks," Journal of Marketing (43), pp. 549–563.

Keller, K. 1993. "Conceptualizing, measuring, and managing customer-based brand equity," Journal of Marketing (57), pp. 1–22.

Keller, K. L. 2003. "Brand synthesis: The multidimensionality of brand knowledge," Journal of Consumer Research (29:4), The Oxford University Press, pp. 595–600.

Keller, K. L., and Lehmann, D. R. 2006. "Brands and Branding: Research Findings and Future Priorities," Marketing Science (25:6), pp. 740–759 (doi: 10.1287/mksc.1050.0153).

King, R. A., Racherla, P., and Bush, V. D. 2014. "What We Know and Don't Know About Online Word-of-Mouth: A Review and Synthesis of the Literature," Journal of Interactive Marketing (28:3), pp. 167–183.

Kleining, G. 1959. "Zum gegenwärtigen Stand der Imageforschung,," Deutschland.

Kohlschütter, C., Fankhauser, P., and Nejdl, W. 2010. "Boilerplate detection using shallow text features,," Presented at the Proceedings of the third ACM international conference on Web search and data mining, ACM, pp. 441–450.

Krishnan, H. S. 1996. "Characteristics of memory associations: A consumer-based brand equity perspective," International Journal of Research in Marketing (13:4), pp. 389–405.

Lawson, R. 1998. "Consumer Knowledge Structures: Networks and Frames," NA - Advances in Consumer Research Volume 25.

Leake, D. B., Maguitman, A., Reichherzer, T., Cañas, A. J., Carvalho, M., Arguedas, M., and Eskridge, T. 2004. "Googling from a concept map: Towards automatic concept-map-based query formation," (Vol. 1), Presented at the Concept maps: Theory, methodology, technology. Proceedings of the first international conference on concept mapping, pp. 409–416.

Liu, B. 2014. "Sentiment Analysis and Opinion Mining," Computational Linguistics, pp. 1–5 (doi: 10.1561/1500000011).

Liu, B., Hu, M., and (null). 2005. "Opinion Observer: Analysing and Comparing Opinions on the Web,," Presented at the Proceedings of the 14th international conference on World Wide Web, Chiba, Japan, April 1, pp. 342–351.

Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M. 2012. "Mine Your Own Business: Market-Structure Surveillance Through Text Mining," Marketing Science (31:3), pp. 521–543 (doi: 10.1287/mksc.1120.0713).

Novak, J. D., and Cañas, A. J. 2008. "The theory underlying concept maps and how to construct and use them,," Institute for Human and Machine Cognition.

Novak, J. D., and Gowin, D. B. 1984. Learning how to learn, Cambridge University Press.

Olson, J. C., and Muderrisoglu, A. 1979. "The Stability of Responses Obtained By Free Elicitation: Implications For Measuring Attribute Salience and Memory Structure," Advances in Consumer Research (6), pp. 269–275.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," Journal of Management Information Systems (24:3), M.E. Sharpe, Inc., pp. 45–77.

Popescu, A., and Etzioni, O. 2005. "Extracting product features and opinions from reviews,," Presented at the Proceedings of the Conference on Human Language Technology and Empirical Methods in Language Processing, pp. 339–346.

Ramos, J. 2003. "Using TF-IDF to Determine Word Relevance in Document Queries,," Presented at the Proceedings of the First Instructional Conference on … (available at http://www.cs.rutgers.edu/mlittman/courses/ml03/ iCML03/papers/ramos.pdf).

Schnittka, O., Sattler, H., and Zenker, S. 2012. "Advanced brand concept maps: A new approach for evaluating the favorability of brand association networks," International Journal of Research in Marketing (29:3), pp. 265–274.

Scott, J. 2012. Social network analysis, Sage.

Supphellen, M. 2000. "Understanding core brand equity: guidelines for in-depth elicitation of brand associations," International Journal of Market Research (42:3), Market Research Society, p. 319.

Teichert, T. A., and Schöntag, K. 2010. "Exploring consumer knowledge structures using associative network analysis," Psychology & Marketing (27:4), Wiley Online Library, pp. 369–398.

Theobald, M., Siddharth, J., and Paepcke, A. 2008. "SpotSigs: robust and efficient near duplicate detection in large web collections,," Presented at the Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, New York: ACM, July 20, pp. 563–570 (doi: 10.1145/1390334.1390431).

Urban, G. L., and Hauser, J. R. 2004. "'Listening in' to find and explore new combinations of customer needs," Journal of Marketing (68:2), American Marketing Association, pp. 72–87.

Xiong, G., and Bharadwaj, S. 2014. "Prerelease buzz evolution patterns and new product performance," Marketing Science (33:3), INFORMS, pp. 401–421.