DETECTING HERDING BEHAVIOR USING TOPIC MINING: THE CASE OF FINANCIAL ANALYSTS

Research paper

Palmer, Matthias, University of Goettingen, Goettingen, Germany, matthias.palmer@uni-goettingen.de

Eickhoff, Matthias, PwC Germany, Business Analytics, Duesseldorf, Germany, matthias.eickhoff@pwc.com

Muntermann, Jan, University of Goettingen, Goettingen, Germany, muntermann@wiwi.uni-goettingen.de

Abstract

The phenomenon of herding behavior can be observed throughout human societies and has been discussed from different perspectives in research areas such as information systems or finance. In this context, financial analysts have been found to exhibit a disposition towards herding behavior. We identify the topic selection within analyst reports as a potential source for the detection of herding behavior among financial analysts. In this paper, we utilize a qualitative dataset of about 140,000 analyst reports and 1,500 conference call transcripts. Building on the topic modeling technique latent Dirichlet allocation, we calculate similarity scores between the respective documents. Potential herding behavior is observable by revealing exceptional topic structures within groups of analyst reports. In this study, these are found especially during a phase of economic recession: the financial crisis spanning from 2008 to 2010. Viewing prospects, this approach might complement existing research in herding behavior by quantifying qualitative data not only within the financial domain but also in other research areas.

Keywords: Herding Behavior, Financial Analysts, Topic Mining, Analyst Reports.

1 Introduction

For financial markets, herding behavior is not just a relevant issue since the advent of the Internet and its wealth of information. Following the actions of others instead of relying on personal information characterizes this effect in the context of economic decision making (Bikhchandani and Sharma, 2001; Hirshleifer and Hong Teoh, 2003). Also, the decision-making process of financial analysts, who play an important role in the financial markets, is affected by a growing amount of information due to digitization. It is the task of these analysts to concisely and transparently filter relevant information within extensive amounts of available data that are not comprehensible for individual market participants without substantial effort (Clement and Tse, 2005). The usage of modern information systems simplifies access to relevant information, but the quantities of available information can be seen as both an opportunity and a challenge (Chen et al., 2012).

In accounting literature, the work of financial analysts has already been associated with and criticized regarding the phenomenon of herding behavior (Youssef and Rajhi, 2009). Herding behavior has been identified within analyst recommendations, which are released through various forms, such as analyst reports or structured recommendations directly available through systems like the Institutional Broker Estimate System (I/B/E/S) (Jegadeesh and Kim, 2010; Trueman, 1994). We see an important step in the analysis of recommendations and price estimates but believe that a detection of herding behavior should be carried out via a large-scale analysis of qualitative data such as analyst reports to find out what is being written about. These analyst reports are written analyst opinions released on an ad-hoc basis, often relating to telephone conference calls of companies. These conference calls take place between analysts and company executives; and usually accompany quarterly earnings releases. Specifically, the topic selection within these sources of information can serve as a tool to identify herding behavior: Analyst reports that have a particularly unusual topic similar to their proximate reports can be used as an indicator for the detection of analysts who simply plagiarize the texts and ideas of their colleagues.

With this idea, we are following up on the latest empirical research that uses network models to uncover herding behavior among financial analysts (Zhao et al., 2014). However, content analysis of analyst reports based on text mining methods does not take place here but in other recently published studies that use these methods to assess the informational and predictive value of analyst reports (Huang et al., 2014). Building on this, there is also relevant research on where analysts primarily obtain their information from (Eickhoff and Muntermann, 2016) and what role they play as intermediaries in revealing information that goes beyond what companies disclose (Huang et al., 2017).

In this paper, we analyze analyst reports and conference calls of all companies listed in the Dow Jones Industrial Average (DJIA) during the years from 2003 to 2015. We include conferences in the analysis to filter for topics that naturally play an important role in analyst reports and are not conducive to our analysis. By using a sufficiently large dataset of about 140,000 reports, we try to minimize the impact of individual events. From a technical perspective, the identification of topics within the dataset is accomplished by using a topic modeling technique called latent Dirichlet allocation (LDA) (Blei et al., 2003); while the similarity between documents is calculated by utilizing the concept of cosine similarity. Using these methods, we combine studies that have already successfully used topic mining for analyst reports in other contexts (Huang et al., 2017; Eickhoff and Muntermann, 2016). A first analysis of the dataset gives an impression of a constant level of similarity between reports. Building on this, we focus our analysis on different time frames resulting from different economic market environments, where an especially strong changing level of similarities between reports supports the detection of herding behavior. We find that the phenomenon of herding behavior distinctively characterizes a phase of economic recession: the financial crisis with its peak in 2008 and the years following until 2010.

This paper is structured as follows: Firstly, an introduction to the work of financial analysts and their potential herding behavior. We follow with a description of the dataset. The next section starts with data pre-processing and gives both a more in-depth definition of LDA and an introduction to the concept of cosine similarity. Subsequently, we describe our analytical approach in detail and present the results of our analysis. Building on these results, the paper follows with a discussion. The final section concludes the paper and provides suggestions for future research.

2 Theoretical Background

2.1 Financial Analysts

Financial analysts face increasing amounts of qualitative sources of information such as news media, company announcements, and social media content. So-called sell-side analysts mainly write reports to give information and purchase recommendations to investors. These analysts cover a certain number of companies – often within specific industries – and provide information to their customers on a regular basis. Furthermore, analysts participate in conference calls in which they can get directly in touch with relevant company representatives. They gather first-hand information on particularly interesting topics, making this medium of communication an important tool of their work and source of analyst opinion.

Analyst reports are texts created by financial analysts to give an overall summary of the condition of companies in the current market environment. On one hand, these reports present the financial situation of a company. On the other hand, they discuss recent topics as well as upcoming opportunities and challenges. Finally, these reports aim to give a purchase recommendation (buy/hold/sell) regarding the analyzed company considering various kinds of relevant information (Womack, 1996): Firstly, financial analysts can take advantage of fundamental analysis, relying on certain business data such as financial statement data or financial ratios. Secondly, recommendations can be based on the overall macroeconomic environment of firms. Thirdly, technical analysis can be of use for analysts. In this case, historical capital market data serves as a basis for their valuation. Furthermore, analysts can appraise the microeconomic environment of the firm and give investment advice based on such contextualization. Most recently, analysts can cover social media content that might be of relevance to the future perspectives of a firm. Most importantly, the abundance of these different possible sources in analyst reports is a broad basis for decision-making for capital market participants and can potentially improve information efficiency (Barber et al., 2001). Above all, the qualitative part of analyst reports is considered helpful in explaining investors' buying recommendations (Huang et al., 2014). In contrast, other research indicates that the market does not necessarily follow the recommendations of analysts directly and without restrictions (Elgers et al., 2001; Bradshaw, 2009; Barniv et al., 2009). However, analyst reports represent a major source in information procurement by investors and are therefore worth to be analyzed.

Conference calls between companies and financial analysts gained substantial importance during the 1990s (Bowen et al., 2002). Primarily, they arose as an attempt towards the mitigation of a dissatisfying low quality in accounting standards (Frankel et al., 1999). These quarterly conferences take place with one of these conferences discussing the annual financial results. Aside from that, conference calls are held in the course of exceptional events, such as upcoming mergers or acquisitions (Crawford Camiciottoli, 2010). Conference calls generally separate into two different parts: The first part contains the companies' presentations, which include the current financial situation as well as corresponding comments. Additionally, this part discusses recent company-specific events and evaluates future development trends. The second part provides an opportunity for analysts to challenge the attending company representatives with their questions. Thus, this two-fold purpose of conference calls allows analysts to get in touch with company representatives to contribute to a more effective information policy of the companies and obtain insights about their future strategies (Tasker, 1998; Crawford Camiciottoli, 2010).

2.2 Herding Behavior among Financial Analysts

Herding behavior in the context of analyst opinions describes the tendency of analysts to avoid bold predictions that contradict the consensus among their peers (Devenow and Welch, 1996). Thus, analysts try to seek safety by "sticking to the herd" (Scharfstein and Stein, 1990). Especially for career reasons, they might be hesitant to give a negative recommendation towards a firm if all their peers praise the same company, even if their own analysis contradicts this general opinion (Hong et al., 2000; Twedt and Rees, 2012). Single analysts tend to assume that information in capital markets is not evenly distributed, meaning that they believe other analysts to have an informational advantage. With regard to that, they adopt assessments conducted by other analysts (Bikhchandani et al., 1992).

Furthermore, increasing amounts of data present a major challenge to financial analysts (Eppler and Mengis, 2004; Simon, 1976). These increasing amounts of information that financial analysts have to process – often in a too short amount of time – has also been discussed as threatening to overwhelm analysts by creating an information overload that might lead to herding behavior (Edmunds and Morris, 2000; Whittaker and Sidner, 1996). This difficulty incites analysts to turn away from a detailed study of all information for limiting and processing only supposedly relevant information (McQueary et al., 2016). This is expressed not only in concrete price recommendations but also in the creation of analyst reports and most notable in the case of analysts not being able to assess the importance of recent topics. Since a certain level of similarity between topics in analyst reports seems reasonable, we therefore suspect that, depending on market environments, this issue might result in a highly increased orientation of individual reports on the topic selection of other analyst reports. Consequently, the following research question emerges*: Does comparing topic similarity in analyst reports for different macroeconomic environments give evidence of herding behavior among financial analysts?*

3 Dataset Description

We use a database containing analyst reports and conference call transcripts for 30 companies listed in the DJIA in the years 2000 to 2015 obtained from Thomson Reuters Advanced Analytics (TRAA). We chose the DJIA instead of another index that includes a higher number of companies as we want to focus on especially influential companies that will be watched by as many analysts as possible. Due to an increase in available conference call transcripts over the course of time, we chose the most recent composition of the companies contained in the DJIA. Overall, 140,196 analyst reports and 1,586 conference call transcripts were collected, resulting in 53 calls and 4,673 reports on average per company. Table 1 lists all companies included in the dataset with the number of analyst reports and conference calls as well as the corresponding industry, which will be referred to in the subsequent analysis.

Industry	Company	Reports	Calls	Industry	Company	Reports	Calls
Oil and Gas	Chevron	3332	55	Con-sumer Services	Disney	4440	56
	Exxon	2991	55		Home Depot	4134	56
Basic Materials	DuPont	3639	56		McDonald's	5314	56
Industri-als	3M	4670	57		Walmart	5531	56
	Boeing	5377	55	Telecom-munica-tions	AT&T	5593	45
	Caterpillar	4034	55		Verizon	5098	55
	General Electric	6196	55	Finan-cials	American Express	3605	55
	United Technologies	4175	55		Goldman Sachs	2857	54
Con-sumer Goods	Coca Cola	3803	56		JPMorgan Chase	4660	57
	Nike	2299	55		Visa	1996	31
	Procter & Gamble	4119	56		Travelers Companies	2242	47
Health Care	Johnson & Johnson	6238	40	Tech-nology	Cisco Systems	6442	52
	Merck	8517	56		IBM	4630	58
	Pfizer	6354	53		Intel	7464	57
	UnitedHealth	4251	35		Microsoft	6195	57

Table 1. Companies included in the dataset with the number of analyst reports and conference calls. The company classification by industry is determined by the Industry Classification Benchmark (ICB, 2017).

A topic comparison between analyst reports is only feasible with a constant amount of reports being published for every individual company in the index. This, in turn, leads to a restriction concerning the dataset due to the availability of data within the chosen index composition. Figure 1 displays the average number of published reports per company in relation to their nearest conference call, which makes up an important part of the subsequent analysis. We chose a time frame of 40 trading days before and after each conference call. With an average of 60 trading days between conference calls, this heatmap displays all reports in the dataset with a theoretical overlap for the days 30 to 40.

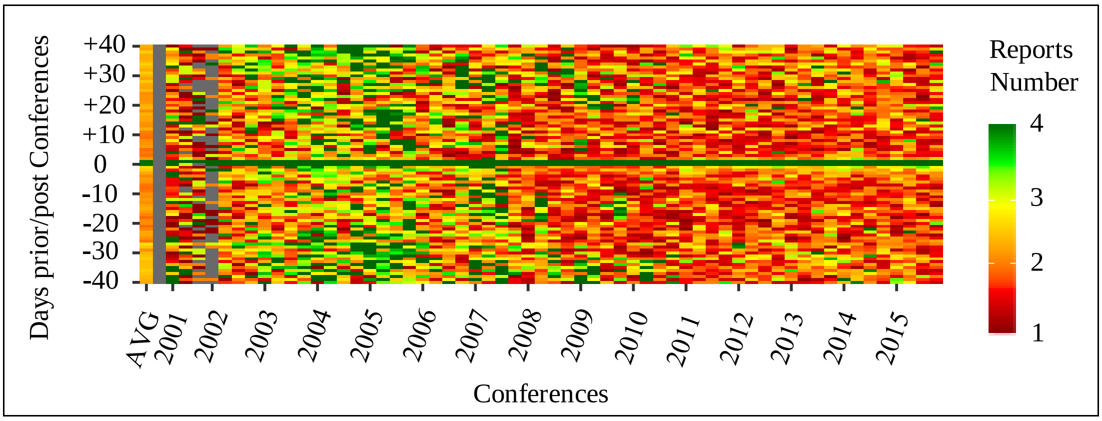

Figure 1. DJIA: Per company average number of published analyst reports. Day 0 represents the day of a conference call in every quarter. For illustration, the maximum values are restricted to the 90%-quantile of the dataset.

As Figure 1 shows, there is a varying number of analyst reports published on average per company in the dataset. Different lengths of availability for analyst reports account for the variation on one hand. For this analysis, we did not search for predecessor companies in the case of a change of the index composition. For example, data corresponding to Visa is available only from mid-2007 onwards. On the other hand, varying phases in the economic cycle cause the level of published reports to change. Furthermore, the heatmap points up to the day of the conference calls and the day after, where analysts seem to release an outstanding number of reports, which is consistent with the results of other analyses in this context (Huang et al., 2017). This phenomenon is observable even though we pruned the heatmap to its 90% quantile to allow for an improved interpretability.

Looking at the years prior to 2003, conference call data is not available for every quarter and the number of analyst reports per conference is relatively low. The period including the years from 2003 to 2007 in contrast to the years from 2008 to 2015 contains a comparatively large number of published analyst reports. Nevertheless, to fulfill the requirement of an analysis dealing with up-to-date data, the period ranging from the years 2008 to 2015 is essential. Specifically, this time frame is useful for the intended comparison of different time frames and macroeconomic environments, respectively. For these reasons, the following analysis utilizes conference calls and analyst reports in the 2003 to 2015 interval.

4 Methodological Approach

4.1 Data Pre-Processing

To ensure a purposive determination of relevant topics, the preparation of the dataset is a crucial step. Specifically, we want to remove figures, tables, and numbers from the dataset. The same holds true for the disclaimer of the documents, otherwise categorized as an independent topic. To identify topics contained in both analyst reports and conference calls, we merge all documents into a single text corpus. We split the pre-processing into three different parts: feature extraction, feature selection, and feature representation (Manning et al., 2008). The following steps have been conducted with the help of the text mining package *tm* in R (Meyer et al., 2008).

Feature extraction: This step aims to convert the text into a representation of single words or phrases. First, we transform the text to lower cases. Next, we filter the text by allowing ASCII characters and numbers only. After the removal of excess whitespace, the text is ready to get split into different tokens. In our case, we tokenize for single words. In the following step, we exclude supposed stopwords from the dataset. These words have no connection to possible topics and occur with a high frequency. To perform this task, we made use of a list of stopwords especially created for English texts. For simplicity, we did not replace any hyponyms by common terms or substitute synonyms. Furthermore, stemming is a common technique to reduce the complexity of a dataset by cutting words back to their common stem (Porter, 1980). Reducing the complexity comes along with an increased computational performance (Witten et al., 1999). However, for the purpose of our paper, we aim at identifying unique topics to help us to compare the similarities between reports. The data that is not stemmed results in an increased granularity of the topic mixture estimated by the topic model due to an increase in the analyzed feature space; which explains why we do not perform stemming in our pre-processing.

Feature selection: The tokenized dataset still consists of many unwanted words or text blocks that negatively impact successful topic mining. We first filter for tokens of a length between a minimum of three and a maximum of 20 characters. In addition, the two percent rarest and the ten percent most frequent words are cut out. After a manual review of the texts, we chose two instead of ten percent for the rarest words as we want to prevent the topics from becoming too unspecific. Filtering by terms that are specific for conference calls, e.g., *quarter*, *dividend*, *year-over-year*, and the name of sell-side analyst companies, as well as the name of the company to be analyzed brings us one step further to a unique topic representation that focusses on higher level conceptual topics, instead of topics that describe which company is being talked about. As a final step, we identify text blocks, which appear several times in the dataset. For example, these text blocks may be part of the disclaimer of the reports and conference calls. To delete these blocks, we perform a separate tokenization by using n-grams that represent text blocks of five words each. If the same text block appears more than five times in the dataset, we delete it because we assume this block to have no special meaning for text understanding.

Feature representation: After cleaning and filtering the dataset, the final part of the pre-processing prepares the dataset for topic mining. Therefore, by the conversion of our dataset into a vector representation (Salton et al., 1975), a matrix known as term-document matrix (TDM) results. This matrix represents every single document as a vector of every word within the document collection. To create more meaningful topics based on text, the TDM can be adjusted by terms occurring either very rarely or frequently. As this adjustment is not necessary for the topic mining algorithm used in this paper, which aims at using as unique as possible topics, we skip this step (Blei et al., 2003).

4.2 Topic Mining and Latent Dirichlet Allocation

The method of topic mining can be considered as a specific form of text mining. Text mining methods strive to process weakly structured textual data. These methods can be used to automatically identify relevant structures or main statements in texts (Hotho et al., 2005). The approach, originally established under the designation “Knowledge Discovery from Text,” is comparable to data mining methods, which collect and analyze quantitative data (Feldman and Dagan, 1995). Topic mining algorithms have been developed against this background (Chen and Liu, 2014). These methods, initially developed to serve as an instrument for dealing with scientific literature, are especially transferable to business applications (Blei and Lafferty, 2009).

Topic models allow us to aggregate information contained in a dataset to be represented by a given number of topics. This leads to a faster process of understanding the content of the documents. Topic identification depends on hyperparameters, which have to be set before estimating topic mining models. For instance, these models depend on the length of an estimation session as well as on its scaling, expressed in the desired number of topics and the number of words per topic (Alghamdi and Alfalqi, 2015). Choosing an appropriate number of topics to be determined by the model can be crucial for the creation of topics that simultaneously represent the underlying topic distribution of the document collection well, while also being interpretable by human inspection. A selected number of topics that is too small might cause relevant topics not being identified. Otherwise, too many topics can lead to the point where it is not possible to assign topics to documents in a reasonable way due to the occurrence of exceedingly specific topics, which would be very similar to human eyes.

Although other approaches, such as bag of words (Harris, 1954) or prediction-based word embeddings (Mikolov et al., 2013), may be more appropriate in certain application scenarios (Dai et al., 2015), LDA is often a preferred approach due to its good interpretability in disciplines such as information systems or management research (Eickhoff and Neuss, 2017). LDA enables us to model the underlying latent topic distribution in the dataset by using a Dirichlet distribution (Blei et al., 2003). This probability model allows us to discuss documents as a mixture of various topics, which are in turn composed of all words included in these documents. In comparison, latent semantic analysis (LSA) decomposes a text corpus by applying singular value decomposition (SVD) and thus transforms the most important words onto a subspace (Dewester et al., 1990). As a result, single words that might have been important for our approach of comparing topics in documents might get lost. Although probabilistic latent semantic analysis (pLSA) might fix this issue by assuming that topics are a distribution of the vocabulary of a document corpus – resulting in non-orthogonal topics –, the approach has strong overfitting problems as the number of parameters strongly increases with an increasing number of texts. When using the existing dataset, we therefore consider this method unsuitable.

Besides a determination of the number of topics for LDA, one must define the Dirichlet parameters α and β, which determine the document-to-topic and the word-to-topic composition, respectively. For example, a high α will lead to the creation of a representation that is characterized by most of the predefined number of topics instead of containing a few highly dominant topics, i.e., α determines the mixture of topics estimated by the model. In turn, a low β creates topics consisting of a small number of relatively dominant words, i.e., β controls the shape of the word-to-topic distribution assumed by the model. Vectors represent the individual topics and every word connected to these topics. Thus, it is possible to calculate the assignment probability with which these words are included in one of the topics. As a result, one can calculate the probability level of every topic with respect to a single document. Conducting these steps, the model creates two matrices: Matrix A contains the probability for words included in single topics and matrix B presents the probabilities for topics to be contained in a document.

This allows for a comparison of different documents and their assignment probabilities for every topic. An exact topic specification is feasible with the help of the probabilities of matrix A. By ranking the word importance within each topic, the most important words qualify as a shorthand title for each individual topic. As an example, Table 2 shows an excerpt of a type-A matrix for a pharma company, Merck, which is included in the upcoming analysis. In this example, we select five exemplary topics from 50 calculated topics to show the variety of topics contained in the document corpus. The assignments of topic titles are as follows: topic 1, Suvorexant (a medication against insomnia); topic 11: portfolio strategy; topic 38: joint ventures and restructuring; topic 45: mergers and acquisitions; topic 50: company outlook. Topics having the highest probabilities in the dataset show an explicit relation to conference calls and reports as the example in Table 2 depicts. On one hand, this illustrates that it is possible to differentiate between topics for the goal of identifying herding behavior by reports' topics. On the other hand, this example illustrates the limits of topic separation being visible in the top 20 words. In this paper, we use the R text mining framework *text2vec* to determine topic probabilities, which provides an implementation of the LDA model described above (Selivanov, 2016).

Topic	Top 20 words per topic
1	placebo, suvorexant, sleep, insomnia, day, baseline, effects, doses, pbo, improvement, min, mean, significantly, receiving, randomized, onset, ambient, effect, statistical, showed
11	portfolio, strategy, focus, commercial, turn, focused, many, areas, brands, morning, united, programs, position, states, moving, take, success, team, pleased, make
38	joint, venture, rose, fell, lined, though, charges, sector, items, foreign, restructuring, ago, lyons, banyu, hilliard, world, johnson, recently, interest, expiration
45	merger, transaction, jnj, combined, rights, agreement, arbitration, Johnson, control, iss, shareholders, animal, settlement, closing, offer, merial, majority, remicadesimponi, pay, remicade
50	upside, case, roughly, outlook, north, much, overweight, though, news, seems, neutral, appears, meaningful, past, nearterm, modest, today, clearly, think

Table 2. Example of a type-A (word-to-topic) matrix for the company Merck. The column "Top 20 words per topic" displays the ranking of words by decreasing probability.

4.3 Cosine Similarity

Building on the LDA method, we need an additional approach to compare the topics contained in different documents. This comparison will be conducted by the calculation of a similarity score between documents. This score can be determined by using the topic assignment similarity vectors of each document and calculating the cosine of the angle between the documents' topic assignment similarity vectors (Han et al., 2011). This so-called cosine similarity is often used in accounting (Loughran and McDonald, 2016) and finance (Hoberg and Phillips, 2010) literature, which analyzes the similarity of texts that are comparable to analyst reports, such as MD&A disclosures (Brown and Tucker, 2011) or annual reports (Lang and Stice-Lawrence, 2015).

For cosine similarity, a higher similarity score between two vectors means that a higher general similarity between the two documents with regard to the covered topics is assumable. The resulting value is bounded [0, 1] for topic vectors, which by default are equal or greater than zero. A high similarity between two documents is expressed by a high cosine value as this is more intuitive than calculating the angle between two vectors. For two vectors, V1 and V2, the cosine similarity is defined as follows:

$$\text{Cosine Similarity} = \frac{\text{Cross product}(V_1, V_2)}{\sqrt{\text{Cross product}(V_1) * \text{Cross product}(V_2)}} \tag{1}$$

5 Analysis

5.1 Analytical Approach

In principle, the analysis consists of two steps: In the first step, the topics contained in each analyst report are estimated using LDA. In the second step, these posterior topic probabilities are utilized to measure the cosine similarity between the analyst reports to uncover possible herding behavior. Despite this basically simple approach, several points must be considered, which are discussed below.

Topic models are heavily reliant on the size and structure of the dataset for which we want to estimate them. For this reason, we pay particular attention to the choice of the composition of our dataset. We could choose between the estimation of topics for a whole stock index, different industries or on company level. A company-specific determination of topics – which means that the same topics are applicable to a single company only – can be justified for the following reasons:

- First, while topics based on whole industries or an index are applicable more flexible than topics depending on a single company, company-specific topics promise to be a more precise tool to measure the similarity between analyst reports. Table 1 lists all analyzed companies and illustrates that even a classification by industry would lead to a topic determination for groups of relatively different companies, as the example of the industry "Consumer Services" suggests.
- Second, reports based on or plagiarizing other reports are most likely to refer to the same company.
- Third, if we allow for every quarter in the observation period one topic to be of high importance for every company, the number of topics for an index or an industry of several companies would be too high for a sensible use of a topic mining algorithm.
- Fourth, the time frame of the dataset should provide a sufficient number of reports per company to estimate a topic model. Table 1 shows that the minimum number of reports for a single company is slightly less than 2000 reports.

Furthermore, we put attention on generating meaningful topics. A manual variation of the predefined number of topics in the LDA model results in choosing 50 topics instead of e.g. 25 or 100. The use of elbow curves would be a useful solution in comparable problems of machine learning, but they are unsuitable for determining the number of topics in LDA (Feuerriegel et al., 2016). We believe that the number of topics is reasonable for the observation period, as this number is roughly based on the number of quarters during the observation period, which is characterized by quarterly reports and conferences.

The previous step was performed each time by varying the LDA parameter α until a satisfying topic composition was observable. Literature suggests a ratio of $50/K$ for the parameter α, where K accounts for the number of topics (Griffiths and Steyvers, 2004). We performed topic calculations for an α of 1, 0.5, 0.1, and 0.01. A manual check of the resulting topics motivated us to differ from the recommendation and use a value of 0.1. This is logical, since we are interested in topics that are even more granular than topics used in standard approaches. The possibility of this manual topic check is the main reason for choosing the LDA algorithm over other approaches briefly addressed in the methods section.

However, the overall topic representation on company level might exhibit single topics that are part of many topic assignment vectors and vice versa. To address this issue, we make use of the idea of the term frequency-inverse document frequency (*Tf-idf*) (Sparck Jones, 1972) with

$$Tf\text{–}idf\ (w) = tf\ \log\left(\frac{N}{df(w)}\right) \tag{2}$$

where w denotes the feature (topic), N the total number of documents, df the term frequency of a feature (topic) in a document, and Tf the term frequency (topic frequency) depending on the respective document (Manning et al., 2008). We transfer the concept to the topic level and adjust the representation of topics with regard to their overall appearance in the dataset of a company. This makes it easier to compare the similarities between the reports of individual days with the total dataset.

As analysts publish reports to a large extent in the time around conference calls, it is likely the case that high cosine similarities emerge from new information that are disclosed during conference calls (Bowen et al., 2002). That is why conference calls are an important part of our analysis, even though we basically want to examine the content of analyst reports. For the purpose of minimizing this effect as much as possible, we will exclude those topics that are mainly discussed in the conferences from the analysis of the analyst reports which are published in the time period after the specific conference calls. In concrete terms, we compare the analyst reports' topic vectors with the topic vectors of the conference calls. Specifically, we subtract the conference calls' topic vectors from the reports' topic vectors. By following this approach, we restrict topic probabilities to have a minimum value of 0. Without the subtraction, the overall level of cosine similarity would be far too high and a meaningful comparison of document similarities over time would be impossible.

Nevertheless, in this analysis, the day at which conference calls take place remains to be an essential building block for the analysis of analyst reports. This particular day, denoted as day 0, will not be part of the analysis. Although we are already attempting to control for the topics of the corresponding conferences, we assume reports published on day 0 are too much influenced by the topic selection in conference calls and will, therefore, lead to a possible misinterpretation of the topic similarity to other reports. Connecting to the last aspect, we chose this approach with reference to other authors who were specifically interested in analyst reports that were influenced by conference calls (Chen et al., 2010). Since we want to avoid this very effect in our analysis, we exclude precisely the reports of those days that are used in their analytical approach.

With conference calls, which take place on a quarterly basis, the observation period has to take this into account. Usually, 60 trading days separate two conference calls. Using longer observation periods containing two or more conference calls would prevent us from matching a report with one specific conference. In addition, the above-mentioned subtraction of the topics of the conferences from the analyst reports would no longer be meaningful to implement. Moreover, with respect to the problem of information overload, the time span between reports and conferences and the disclosure of new information has to be of an appropriate length. By using a long observation period, it is not possible to address the problem of time-critical information processing, which might lead to herding behavior. With an observation period that is too short, thoroughly created reports would not be considered in the analysis, as they are likely to be released later. Additionally, this observation frame would not consider reports written by well-informed analysts who publish their work far in advance of upcoming conferences while referring to the past conference call. Consequently, we decide to work with an observation period that covers 60 trading days following each conference call available in the dataset.

For the purpose of our analysis, we calculate cosine similarities between every analyst report and the last five analyst reports published at least one day prior to it by sticking to a time frame of trading day 2 until trading day 60 post of each conference call. From a technical perspective, we ensure that we only include those reports in our analysis for which five more reports are available on day 1 or prior to the analyzed trading day. This might lead to an omission of single reports but ensures a meaningful comparison of cosine similarities throughout the whole observation period. The selection of five reports results from the composition of the dataset since with this choice a sufficient number of previous reports can be assigned to each report. Also, we believe analysts are guided by a handful of recent reports, hence the natural choice of the previous five reports seems to be appropriate. In summary, that means that if the average cosine similarity between the primary report and the five previous reports is relatively high, the topic similarity might be an indicator of herding behavior.

In detail, our approach includes the following aspects: We make sure that for each comparison the cosine similarities are calculated for single reports only. That means that we need to do five calculations for each report. Calculating average topic probability vectors for the five preceding reports would result in an upshift of the cosine similarity level since the resulting average vector exhibits matches with a higher number of topics. This will increase the likelihood of a similarity to the analyzed report. Next, we calculate the average of the resulting cosine similarities for all individual reports of a trading day. As our focus lies on an overall understanding of herding behavior, we compute averages for each trading day for all companies contained in our dataset. This average always refers to the respective quarter and the days since the conference call in that specific quarter. Consequently, we can compare average cosine similarities for a specific time frame with the overall level of cosine similarity of the DJIA in the observation period. It is reasonable to assume that similarities between reports are also explained by other events than conference calls. For the following reasons, we think that this limitation can be handled: First, it is not possible to calculate with these events individually. Second, excluding every possible such confounding event would probably not leave any events that cause herding behavior. We think that the largest possible dataset is a suitable means of countering these adversities, i.e., we conduct our analysis at the index level and not for individual companies.

The number of reports plotted in Figure 1 qualifies for the selection of observation periods of different lengths. A shift in the average number of reports at the end of the year 2007 supposes a division of the dataset into two parts consisting of a pre-2008 and a time frame from 2008 to 2015. The importance of an analysis examining cosine similarity over time is also exhibited in Figure 2, which outlines the average trading volume per company for the DJIA. Although this figure seems to be influenced by the impact of high frequency trading, it helps to map trading volume as a possible indicator for different phases within the economic cycle. Trading volumes slightly increase from years 2003 to 2008. Between 2008 and 2010, trading volumes peak; indicating market turbulences, which are of special interest to an analysis of herding behavior. Trading volume is at its peak especially around conference calls. From 2011, trading volume decreases with peaks for single quarters. Principally, neglecting phases with low trading volume might be reasonable. For the time span between 2014 and 2015, one could argue that low trading volumes coincide with a reduced number of reports published as shown in Figure 1. In accordance with the explanation given for the number of reports, we keep these parts of the dataset in the interest of providing an up-to-date analysis. Aside from that, the identified phases within the heat map are potentially indicative of herding behavior as turbulences increase the stress financial analysts must endure. In conclusion, it seems reasonable to split the dataset into three parts. Part one covers the time frame from 2003 until the end of 2007, mostly representing a period of economic growth. Part two represents the financial crisis and its aftermath. Therefore, we define this period as of the beginning of the year 2008 until the end of the year 2010. Part three represents the post-crisis period from the year 2011 to 2015.

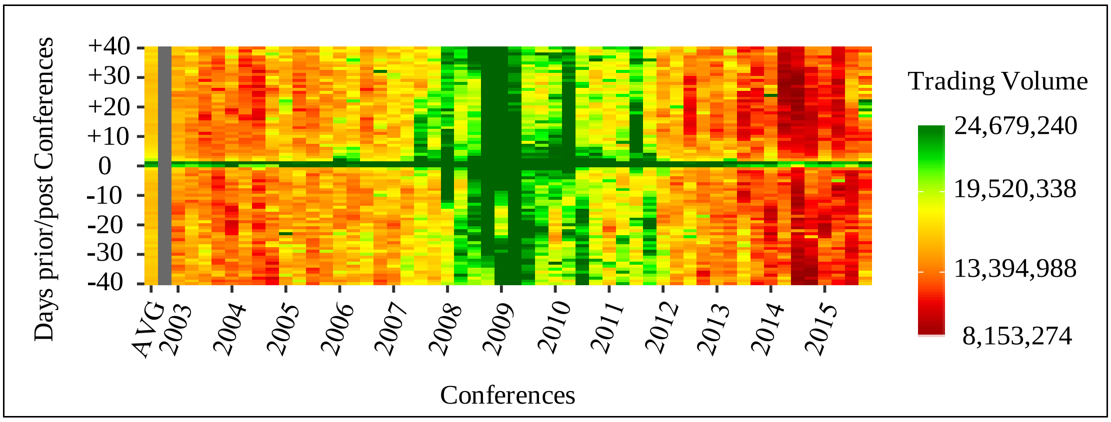

Figure 2. DJIA: Average trading volume per company. For illustration, maximum values are restricted to the 90%-quantile.

5.2 Empirical Results

Figure 3 illustrates the results of the analysis using the entire dataset. For example, the topics that are already listed in Table 2 are taken from this dataset. Cosine similarity values are shown subsequent to day 1 following each conference call. For this day, we determine no results as we excluded the reports of day 0, to which in our analysis the reports of day 1 would be compared to. Figure 3 indicates an overall constant level of cosine similarity of 0.17 between reports. The value depends in particular on the reports themselves and the number of topics used in the analysis. A comparison with other text types does not seem sensible since they cannot be compared with the structures of analyst reports and the natural similarities resulting from that. To our knowledge, there are no comparable studies with which we could compare this value. This does not pose a problem to our analysis, since we are only interested in an indication of a level of similarity in this part of the analysis.

Although we cannot detect abnormal levels of cosine similarity in any specific time frame after the conference calls, Figure 3 proves the success of eliminating major conference call topics in our analyzed reports. Otherwise, there should be a peak of cosine similarity immediately following the conference calls. Nevertheless, Figure 3 gives an indication of the standard constant similarity level between reports, which cannot be eliminated by our pre-processing techniques nor by an improvement of the topic assignments. These results form the basis for a more detailed examination of the dataset, in which the different time frames already outlined in section 5.1 can be considered.

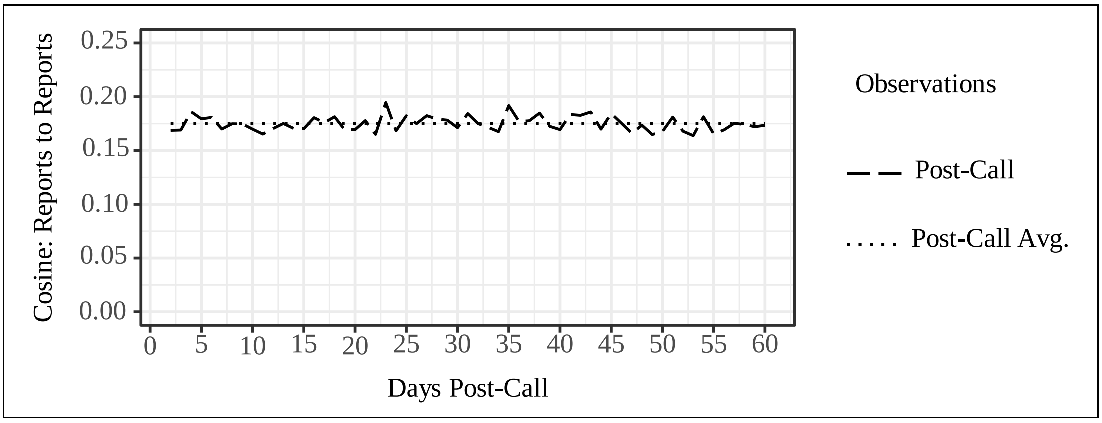

Figure 3. *Average cosine similarity between reports as calculated by the outlined analytical approach. The dataset used for this plot consists of all companies in the utilized DJIA sample, an LDA model for 50 topics, and an observation period from 2003 to 2015.*

Figure 4 plots the results of our analysis for three different time frames, where we can observe differences between all of them. As follows, the results of each time frame will be analyzed:

The time frame of 2003 to 2007 depicts a higher average cosine similarity of 0.21 than the average for the entire dataset. The overall higher level of cosine similarity in comparison to the other time frames displayed in Figure 4 might indicate for herding behavior. On the other hand, the cosine similarity is much more volatile, but without giving proof for special trends. In addition, in the case of a shift in the level of cosine similarity, we assume that this is due to a less digitized environment in which analysts' information gathering was harder overall and thus a smaller number of topics in reports were discussed. Also, a comparably less volatile market environment could affect that as well. Therefore, we see no signs of herding behavior for this period.

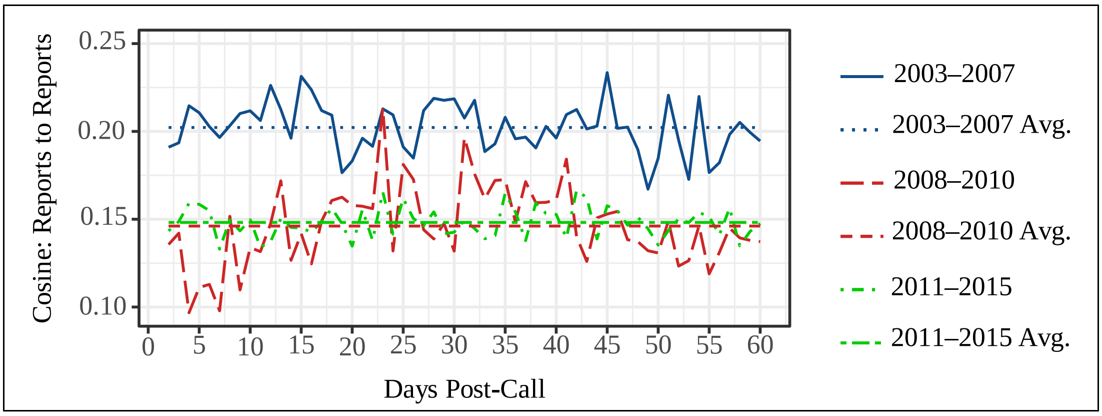

Figure 4. *Average cosine similarity between reports in three different time frames. The y-axis is adjusted for a better comparison of the time frames. The dataset used for this plot consists of all companies in the utilized DJIA sample and an LDA model for 50 topics.*

For the time frame of the years 2008 and 2010, as plotted in Figure 4, the average cosine similarity is on a lower level than for the entire dataset covering the years 2003 to 2015. Due to a complex market environment and a volatile disclosure of new information, analyst reports tend to be less similar on average. However, apart from that, Figure 4 shows an increased cosine similarity for the trading days 20 to 45 after the conference calls. As there is such a big shift in the level of cosine similarity, we see this as an indicator of herding behavior. In times of crisis and when analyst reports are not published directly in relation to conference calls, there seems to be an increasing focus on the reports of other analysts.

Continuing with the period from 2011 to 2015, we denote an average cosine similarity of 0.14. Compared to the other time frames, the graph is less volatile and also cannot verify any patterns that show a systematic change in herding behavior. The average cosine similarity of this period and the period of 2008 and 2010 is nearly on the same level, even though we compare totally different macroeconomic environments. One indicator for this observation and the huge difference in the average cosine similarity to the time frame of 2003 to 2007 might be due to the following reason: Figure 1 shows notable numbers of published reports only for a short time frame after the conference calls. Since we attempt to eliminate the effect of high cosine similarities during that time frame by a comparison to the topic assignment of the corresponding conference calls, we suppose this time frame to be largely affected by these topics. Therefore, the observed average seems to be reasonable.

Pointing towards the research question, we can verify that a topic mining approach allows for a comparison of different levels of cosine similarities, thus enabling the detection of herd behavior. This is mainly achieved by putting forward a possible way to differentiate between natural similarity between reports and herding behavior. Specifically, Figure 4 displays this ability in a meaningful way. We identify indications of herding behavior for a phase of economic recession of the years from 2008 to 2010.

6 Discussion

The findings outlined above can influence both practical and theoretical aspects regarding financial analysts' work. Capital market participants get a sense of when to critically scrutinize the contents of analyst reports. Phases of possible herding behavior can arise when many analysts do not know what topics to talk about. In turn, this phenomenon can be a consequence of an unstable market environment in conjunction with analysts who are suboptimally informed. However, the topics contained in the analyst reports might also reveal whether individual, oftentimes inexperienced (Hong et al., 2000), analysts basically address the right topics but give a price target based on other analyst opinions out of fear of being wrong (Trueman, 1994). In addition, an analysis of topics in analyst reports could also explore how analysts increase the number of reports for career reasons without really adding new content. Also, equity research companies itself can better gauge their analysts' characteristics and consequently distinguish between phases they should possibly adjust their number of employed analysts. For example, the concept of information overload, as discussed by Edmunds and Morris (2000), could be countered by this. In addition, a better overview and a structured comparison of the topics that analysts write about might be beneficial to research in the field of analysts' compensation (Groysberg et al., 2011). Above all, the analysis of similarity between qualitative analyst data can be a supplement to existing research approaches that operate on a purely quantitative basis, e.g., network approaches to detect herding behavior (Zhao et al., 2014).

It must be noted that the presented results only apply to the DJIA and for a limited period between 2003 and 2015. Apart from that, reports published on the same day during which the conference calls are taking place are not included in the analysis, which reduces the dataset. Furthermore, we calculated an average for all individually calculated similarity scores between reports. Strong herding behavior with respect to single topics in single reports might get lost with this kind of approach. Finally, the results depend on topic quality, which is in turn affected by data quality and the used topic mining method. Despite all efforts to separate natural similarity from herding behavior, a concrete distinction will remain difficult. It can never be said exactly whether it is not just expert knowledge or even really new information that affects all analysts equally strongly.

7 Conclusion and Future Research

We utilized the topic mining technique LDA and the concept of cosine similarity to identify herding behavior among financial analysts. Based on a dataset containing companies listed in the DJIA for the years from 2003 to 2015, the results show indications of herding behavior. In particular, this holds true for a phase of economic recession: the financial crisis from 2008 to 2010. From a practical perspective, our approach achieves awareness among capital market participants towards the influence of herding behavior on the content of analyst reports. From a theoretical point of view, our findings underline the existing research dealing with herding behavior and the quantification of qualitative data regarding financial analysts.

For future research, expanding the dataset and the observation period would be the next step towards a verification of the results. Another extension could be the calculation of report-to-report similarities for different numbers of reports and a comparison of similarity with reports that are completely unrelated. In combination with topics estimated for specific periods and their environments, this change in the analytical approach could lead to a better identification of herding behavior. Moreover, dynamic topic models could increase the actuality reflected in topics (Blei and Lafferty, 2006). Building on this, new approaches that combine word embeddings and the interpretability of LDA seem to be promising (Moody, 2016). Furthermore, both different document classification methods and different similarity measures could be used to validate the results of this analysis, especially when it comes to classification methods that are statistically testable. Also, a search for frequently appearing n-grams might detect direct copying of texts. An expansion to additional document types related to the work of analysts might complement the analysis. Further work on analyzing financial analysts' behavior is especially promising for the time frame of the years 2011 until today due to an increased number of analysts covering companies. Finally, a change to a different content domain provides an opportunity to assist in the detection of herding behavior for fields of application characterized by qualitative data, such as marketing.

References

Alghamdi, R. and K. Alfalqi (2015). "A Survey of Topic Modeling in Text Mining." *International Journal of Advanced Computer Science and Applications* 6 (1), 147–153.

Barber, B., R. Lehavy, M. McNichols and B. Trueman (2001). "Can Investors Profit from the Prophets? Security Analyst Recommendations and Stock Returns." *The Journal of Finance* 56 (2), 531–563.

Barniv, R., O.-K. Hope, M. J. Myring and W. B. Thomas (2009). "Do Analysts Practice What They Preach and Should Investors Listen? Effects of Recent Regulations." *The Accounting Review* 84 (4), 1015–1039.

Bikhchandani, S., D. Hirshleifer and I. Welch (1992). "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy* 100 (5), 992–1026.

Bikhchandani, S. and S. Sharma (2001). "Herd Behavior in Financial Markets." *IMF Staff Papers* 47 (3), 279–310.

Blei, D. M. and J. D. Lafferty (2006). "Dynamic Topic Models." In*: Proceedings of the 23rd International Conference on Machine Learning,* Pittsburgh, USA, 113–120.

Blei, D. M. and J. D. Lafferty (2009). Topic Models. In: *Text Mining: Classification, Clustering, and Applications.* Boca Raton: CRC Press, 71–94.

Blei, D. M., A. Y. Ng and M. I. Jordan (2003). "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3, 993–1022.

Bowen, R. M., A. K. Davis and D. A. Matsumoto (2002). "Do Conference Calls Affect Analysts' Forecasts?" *The Accounting Review* 77 (2), 285–316.

Bradshaw, M. T. (2009). "Analyst Information Processing, Financial Regulation, and Academic Research." *The Accounting Review* 84 (4), 1073–1083.

Brown, S. V. and J. W. Tucker (2011). "Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications." *Journal of Accounting Research* 49 (2), 309–346.

Chen, H., R. H. Chiang and V. C. Storey (2012). "Business Intelligence and Analytics: From Big Data to Big Impact." *MIS Quarterly* 36 (4), 1165–1188.

Chen, X., Q. Cheng and K. Lo (2010). "On the Relationship Between Analyst Reports and Corporate Disclosures: Exploring the Roles of Information Discovery and Interpretation." *Journal of Accounting and Economics* 49 (3), 206–226.

Chen, Z. and B. Liu (2014). "Mining Topics in Documents: Standing on the Shoulders of Big Data." In*: Proceedings of the 20th International Conference on Knowledge Discovery and Data Mining,* New York, USA, 1116–1125.

Clement, M. B. and S. Y. Tse (2005). "Financial Analyst Characteristics and Herding Behavior in Forecasting." *The Journal of Finance* 60 (1), 307–341.

Crawford Camiciottoli, B. (2010). "Earnings Calls: Exploring an Emerging Financial Reporting Genre." *Discourse & Communication* 4 (4), 343–359.

Dai, A. M., C. Olah and Q. V. Le (2015). "Document Embedding with Paragraph Vectors." *arXiv preprint arXiv:1507.07998*.

Devenow, A. and I. Welch (1996). "Rational Herding in Financial Economics." *European Economic Review* 40 (3), 603–615.

Edmunds, A. and A. Morris (2000). "The Problem of Information Overload in Business Organisations: A Review of the Literature." *International Journal of Information Management* 20 (1), 17–28.

Eickhoff, M. and J. Muntermann (2016). "They Talk but What do They Listen to? Analyzing Financial Analysts' Information Processing Using Latent Dirichlet Allocation." In*: Proceedings of the 20th Pacific Asia Conference on Information Systems,* Chiayi, Taiwan, paper 312.

Eickhoff, M. and N. Neuss (2017). "Topic Modelling Methodology: Its Use in Information Systems and Other Managerial Disciplines." In*: Proceedings of the 25th European Conference on Information Systems,* Guimarães, Portugal, 1327–1347.

Elgers, P. T., M. H. Lo and R. J. Pfeiffer Jr. (2001). "Delayed Security Price Adjustments to Financial Analysts' Forecasts of Annual Earnings." *The Accounting Review* 76 (4), 613–632.

Eppler, M. J. and J. Mengis (2004). "The Concept of Information Overload: A Review of Literature From Organization Science, Accounting, Marketing, MIS, and Related Disciplines." *The Information Society* 20 (5), 325–344.

Feldman, R. and I. Dagan (1995). "Knowledge Discovery in Textual Databases (KDT)." In*: Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining,* Montréal, Canada, 112–117.

Feuerriegel, S., A. Ratku and D. Neumann (2016). "Analysis of How Underlying Topics in Financial News Affect Stock Prices Using Latent Dirichlet Allocation." In*: Proceedings of the 49th Hawaii International Conference on System Sciences,* Kauai, USA, 1072–1081.

Frankel, R., M. Johnson and D. J. Skinner (1999). "An Empirical Examination of Conference Calls as a Voluntary Disclosure Medium." *Journal of Accounting Research* 37 (1), 133–150.

Griffiths, T. L. and M. Steyvers (2004). "Finding Scientific Topics." *Proceedings of the National Academy of Sciences* 101 (1), 5228–5235.

Groysberg, B., P. M. Healy and D. A. Maber (2011). "What Drives Sell-Side Analyst Compensation at High-Status Investment Banks?" *Journal of Accounting Research* 49 (4), 969–1000.

Han, J., J. Pei and M. Kamber 2011. *Data Mining: Concepts and Techniques (Morgan Kaufmann Series in Data Management Systems).* 3rd Edition. Oxford: Elsevier.

Harris, Z. S. (1954). "Distributional Structure." *Word* 10 (2–3), 146–162.

Hirshleifer, D. and S. Hong Teoh (2003). "Herd Behaviour and Cascading in Capital Markets: A Review and Synthesis." *European Financial Management* 9 (1), 25–66.

Hoberg, G. and G. Phillips (2010). "Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis." *The Review of Financial Studies* 23 (10), 3773–3811.

Hong, H., J. D. Kubik and A. Solomon (2000). "Security Analysts' Career Concerns and Herding of Earnings Forecasts." *The Rand Journal of Economics*, 121–144.

Hotho, A., A. Nürnberger and G. Paaß (2005). "A Brief Survey of Text Mining." In*: Ldv Forum,* 19–62.

Huang, A. H., R. Lehavy, A. Y. Zang and R. Zheng (2017). "Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach." *Management Science*, 1–23.

Huang, A. H., A. Y. Zang and R. Zheng (2014). "Evidence on the Information Content of Text in Analyst Reports." *The Accounting Review* 89 (6), 2151–2180.

ICB. (2017). *Industry Classification Benchmark.* URL: http://www.icbenchmark.com/ICBDocs/FTSE_ICB_Corporate_Brochure.pdf (visited on 4/17/2018).

Jegadeesh, N. and W. Kim (2010). "Do Analysts Herd? An Analysis of Recommendations and Market Reactions." *Review of Financial Studies* 23 (2), 901–937.

Lang, M. and L. Stice-Lawrence (2015). "Textual Analysis and International Financial Reporting: Large Sample Evidence." *Journal of Accounting and Economics* 60 (2), 110–135.

Loughran, T. and B. McDonald (2016). "Textual Analysis in Accounting and Finance: A Survey." *Journal of Accounting Research* 54 (4), 1187–1230.

Manning, C. D., P. Raghavan and H. Schütze 2008. *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press

McQueary, B., A. DeFrancesco, L. Lehman and R. Dziegiel Jr. (2016). "Adaptive Workspace to Reduce Analyst Information Overload." In*: Proceedings of the International Conference on Artificial Intelligence*, Las Vegas, USA, 437–443.

Meyer, D., K. Hornik and I. Feinerer (2008). "Text Mining Infrastructure in R." *Journal of Statistical Software* 25 (5), 1–54.

Mikolov, T., K. Chen, G. Corrado and J. Dean (2013). "Efficient Estimation of Word Representations in Vector Space." *arXiv preprint arXiv:1301.3781*.

Moody, C. E. (2016). "Mixing Dirichlet Topic Models and Word Embeddings to Make lda2vec." *arXiv preprint arXiv:1605.02019*.

Porter, M. F. (1980). "An Algorithm for Suffix Stripping." *Program* 14 (3), 130–137.

Salton, G., A. Wong and C.-S. Yang (1975). "A Vector Space Model for Automatic Indexing." *Communications of the ACM* 18 (11), 613–620.

Scharfstein, D. S. and J. C. Stein (1990). "Herd Behavior and Investment." *The American Economic Review* 80 (3), 465–479.

Selivanov, D. 2016. text2vec: Modern Text Mining Framework for R.

Simon, H. A. 1976. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization.* 3rd Edition. New York: Free Press.

Sparck Jones, K. (1972). "A Statistical Interpretation of Term Specificity and its Application in Retrieval." *Journal of Documentation* 28 (1), 11–21.

Tasker, S. C. (1998). "Bridging the Information Gap: Quarterly Conference Calls as a Medium for Voluntary Disclosure." *Review of Accounting Studies* 3 (1), 137–167.

Trueman, B. (1994). "Analyst Forecasts and Herding Behavior." *The Review of Financial Studies* 7 (1), 97–124.

Twedt, B. and L. Rees (2012). "Reading Between the Lines: An Empirical Examination of Qualitative Attributes of Financial Analysts' Reports." *Journal of Accounting and Public Policy* 31 (1), 1–21.

Whittaker, S. and C. Sidner (1996). "Email Overload: Exploring Personal Information Management of Email." In*: Proceedings of the 14th ACM Conference on Human Factors in Computing Systems,* British Columbia: Canada, 276–283.

Witten, I. H., A. Moffat and T. C. Bell 1999. *Managing Gigabytes: Compressing and Indexing Documents and Images.* 1st Edition. San Francisco: Morgan Kaufmann.

Womack, K. L. (1996). "Do Brokerage Analysts' Recommendations Have Investment Value?" *The Journal of Finance* 51 (1), 137–167.

Youssef, M. and M. T. Rajhi (2009). "Determinants of Herding Behavior Among Financial Analysts: A Study of French Listed Firms." *International Journal of Business and Management* 4 (9), 121–130.

Zhao, Z., Y. Zhang, X. Feng and W. Zhang (2014). "An Analysis of Herding Behavior in Security Analysts' Networks." *Physica A: Statistical Mechanics and its Applications* 413 (1), 116–124.