Summer 6-27-2016

THEY TALK BUT WHAT DO THEY LISTEN TO? ANALYZING FINANCIAL ANALYSTS' INFORMATION PROCESSING USING LATENT DIRICHLET ALLOCATION

Matthias Eickhoff
University of Göttingen, matthias.eickhoff@wiwi.uni-goettingen.de

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

THEY TALK BUT WHAT DO THEY LISTEN TO? ANALYZING FINANCIAL ANALYSTS' INFORMATION PROCESSING USING LATENT DIRICHLET ALLOCATION

Matthias Eickhoff, Chair of Electronic Finance and Digital Markets, University of Göttingen, Germany, Matthias.Eickhoff@wiwi.uni-goettingen.de

Jan Muntermann, Chair of Electronic Finance and Digital Markets, University of Göttingen, Germany, Muntermann@wiwi.uni-goettingen.de

Abstract

In this study, we examine stock analyst information processing behaviour on the example of information transfer between analyst conference calls and analyst reports. From a theoretical perspective, the study contributes to an understanding of analysts' recommendation biases resulting from their information processing. It provides new insights on how information is actually used by analysts, while practical implications for both sides of conference calls and other market participants are examined. Results indicate that analysts are exposed to new information during conference call events, which they consequently incorporate in their reporting.

Keywords: Topic-Mining, Text-Mining, Financial Analysts, Information Processing

1 INTRODUCTION

The rapid increase of available textual information has been discussed in research for a long time (Baker and McCallum 1998). Financial analysts, who we define as professional analysts being paid to provide their analysis, as a group are especially exposed to this increase in amount and velocity of information, as they are, by job definition, expected to provide an analysis that encompasses the available information. Analysts always been facing an overwhelming amount of information as the stock markets have always provided amounts of information that escape an individual's scrutiny. Thus, the group may have developed coping mechanisms that are also suitable for adoption in other domains. However, the information processing of analysts has been criticized regarding a multitude of biases in the past, some of which we will discuss in the following section. Due to this union of capability and (at least partially) understood inefficiency, financial analysts provide an interesting research field regarding information processing behaviour. In order to make this behaviour visible, we employ Latent Dirichlet Allocation (LDA), a modern topic mining algorithm, to two kinds of textual data produced by financial analysts and their communication with company representatives. Using this topic-modelling approach, we study how topics transfer between the two content domains by tracking the topic-similarity around conference call events. This research is structured as follows. In following section, we provide theoretical background regarding financial analysts and their known biases described in literature. In the third section, we describe the two types of content used in our analysis and how we prepare the texts for the following analysis. Following the description of our pre-processing, we provide a short introduction to the applied methodology, which begins by explaining the LDA topic model and introduces cosine similarity as a measure for topic-distance in the context of our analysis. In the fifth section of the paper, we provide the results of our analysis for both a single conference call and the entire sample for varying subsample selections. Finally, we discuss the implications of this research for future research in this domain, as well as implications for concerned practitioners, and the transferability of the chosen approach to other domains. We conclude the paper with a discussion of limitations of the chosen approach and a result summary.

2 THEORETICAL BACKGROUND

Professional financial analysts are known to exhibit a number of inefficiencies regarding their information processing habits. Prior research suggests that this is caused by a multitude of factors, such as herding behaviour, i.e. the tendency to stick to the consensus estimate (Twedt and Rees 2012 p. 2), due to career concerns (Clement and Tse 2005), fear of being singled out in the case of wrong predictions (Hong et al. 2000) or inadequate incentive structures, focusing on increasing brokerage or investment banking revenue instead of rewarding correct predictions (Groysberg et al. 2011). While this prior research focuses on identifying inefficiencies in analysts' information processing, we analyze how analysts actually arrive at their conclusion. Specifically, we focus on the topics they talk about in conference calls and write about in their reports on the basis of which they finally justify their conclusion. Professional analysts are faced by a multitude of possible sources of unstructured information, such as newswire services, corporate disclosures or social media. Obviously, these sources of information are available to anyone truly interested in analyzing a firm's performance. However, analysts may also have access to privileged information. Since decreasing information asymmetry between market participants is a prerequisite for efficient capital markets, fair disclosure regulation aims at reducing the occurrence of privileged communication between firms and analysts. Still, the effect of such regulation may have unintended adverse consequences (Sunder 2002; Irani and Karamanou 2003). One possible channel for such privileged information is presented by analyst telephone conferences in which analysts are able to directly interact with high level, often C-level, representatives of the concerned company. As this degree of direct interaction with high-level employees is unusual, analyst calls are supposed to reveal new information about a company's current and future prospects. Even if the content of the calls is disclosed to the public in a timely manner, the call itself provides analysts with the opportunity to ask company representatives the questions that are important for their specific information needs, thus providing the analyst with information that may not be of value to other market participants. We apply topic-mining techniques to both types of content and examine the topic similarity between analyst conference calls and reports written prior to and after these telephone conferences. This allows us to compare the topic structure of the calls to reports in two directions. First, we examine how related reports that were released prior to the call events are to the topic structure of the calls. This allows us to assess if the topics discussed in the calls are 'new', or if calls mainly discuss topics already addressed in previous reports. This also relates to the question of analyst herding behavior. If analysts do not bring up new topics in the calls, this may be a sign of 'sticking to the herd'. Based on these considerations, we pose the following research questions**:**

(1) To what extend do financial analysts include analyst reports released prior to conference calls in their topic-selection during the call?

If conference calls are sources of valuable new information, the topics discussed therein should be picked up after the call, i.e. the reports following the call should show increased similarity to the call when compared to the ones released prior to the call. Thus, we pose our second research question:

(2) To what extend do financial analysts include conference call topics when writing post-call reports?

The answers to these questions are of both theoretical and practical importance. First, if analysts do not incorporate novel information revealed in analyst calls in their assessments, analyst calls are of little informational value to the analyst and merely a media outlet for companies. Second, if analysts do make use of novel information revealed in calls, it enables market participants to study their content and predict analyst opinions. While the topics discussed during conference calls are influenced by both corporate and analyst participants, the firm holding the call could try to influence the development of the call by adjusting their presentation, which is usually a comparatively long monologue at the beginning of the call. The same holds true for analysts themselves. By gaining insights into what topics have affected analyst behaviour in the past, analysts may be able to recognize manipulation attempts by corporate representatives or improve their own information processing.

3 DATA AND PRE-PROCESSING

For the purposes of this study, we make use of two sources of unstructured data, which we subject to modern text mining methods. For both data sources, we have collected comprehensive data for the multinational technology and consulting corporation IBM between 2000 and 2015. In this period, 59 calls and 4735 reports were collected. Of these 59 calls 48 will be used for the analysis as some calls do not satisfy the conditions required for later analysis, i.e. that they are surrounded by enough analyst reports. Both datasets were downloaded from Thomson Reuters Advanced Analytics (TRAA).

Analyst Earnings Call Transcripts: These transcripts contain the communication between corporate representatives and selected analysts. In the case of IBM, its vice president for investor relations, as well as its CFO typically represent the company, while a small group of analysts (~10) joins them. A typical call consists of an initial presentation by the company, followed by a Q&A section, in which questions by the analysts are answered by the corporate representatives.

Analyst Reports: These reports are available as PDF files from which we extract the textual content needed for our analysis. Furthermore, we identify the date on which reports were added to the TRAA database and the company responsible for their release. Both types of documents are pre-processed by removing 'stopwords', i.e. common English words which are not expected to help us establish the differences between texts due to their omnipresence, and non-textual content, such as tables, figures, or reoccurring numbering (which would otherwise become a 'topic'), before submitting these documents to the topic-mining described in the next section. Also, the disclaimer contained in each call transcript is removed for the same reasons. Consequently, both types of documents are combined into a single textual corpus, stemmed, and converted to a bag-of-words vector space representation, i.e. a term-document-matrix (TDM). We filter the TDM for both sparse and very frequent terms, neither of which are expected to be helpful in establishing statistical differences between the texts, before dropping all documents, which contain no words satisfying these restrictions. We do not convert the TDM to a TF-IDF matrix because of theoretical considerations concerning the chosen topic mining algorithm (Blei et al.,2003).

4 METHOD

Our analysis will be conducted in two steps: First, we address RQ1 by utilizing Latent Dirichlet Allocation (LDA), a topic mining algorithm, on both types of documents and compute similarity scores between the resulting vectors of topic to document probability. If analysts incorporate novel information released in conference calls into their reports, similarity scores between the two types of content should increase if a call took place prior to the release of a report. RQ2 is addressed by interpreting our results regarding their theoretical implications and possible implications for both financial analysts and corporate representatives in analyst calls. Latent Dirichlet Allocation (Blei et al. 2003) generates a statistical model representing the latent topic distribution of a given document collection. The model describes the documents as a mixture of topics (Θ), which in turn are a distribution of the words contained in the documents (N). For both mixtures, a vector of assignment probabilities is calculated. Each word receives a (conditional) per-topic assignment probability and each topic is in turn assigned to each document with another conditional probability. Or formally, following the definition of the generative process by (Blei et al. 2003): A corpus **D** is a vector of documents **w**, each of which is in turn comprised of **N** individual words w_n. Thus, for each document vector **w** in corpus **D** the topic distribution over documents and per-topic word distributions are computed as shown in Table 1.

1. Choose **N** ~ Poisson(ξ), i.e. a word distribution.

2. Choose **Θ** ~ Dirichlet(α), i.e. a topic distribution.

3. For each of the **N** words w_n:

- **I:** Choose topic z_n ~ Multinomial(**Θ**).
- **II:** Choose a word w_n from $p(w_n|z_n, \beta)$, i.e. the multinomial conditional probability of the word conditioned on the topic z_n.

Table 1: Description of the Latent Dirichlet Allocation Topic-Mining Algorithm following (Blei et al. 2003). ***N** refers to the number of words contained in a document, **Θ** denotes the topic distribution.*

This process results in two matrices. Matrix **A** contains the word to topic probabilities, while matrix **B** contains the topic to document probabilities. Matrix **A** can be used to identify the overarching theme of a particular topic by ordering the matrix by descending word probabilities for each topic. Matrix **B** can be used to compare different documents using their (dis)similarity regarding these topics. We use the R 'topicmodels' package for our analysis (Grün and Hornik 2011). When training the topic model, the main parameter that needs to be chosen is k, the number of topics included in the model. The choice of k is a trade-off between choosing a small value, which trains a model with very few topics that are quite distinct from one another, and a large k, which results in a model with many topics, however, these topics may be more similar to each other. Another factor in this choice is the number of topics that are expected to be naturally included in the analyzed content. As discussed, the LDA model provides a matrix of topic to document assignment probabilities (B). This can be used to compare documents in a number of ways. A common approach is the computation of similarity scores between the respective topic probability vectors of two documents. A higher similarity implies a more similar topic structure and consequently similar documents that are more alike. Different measures for these resemblances have been proposed and such similarity scores are not inherently alike human intuition about document similarity (Lee et al. 2005). Here, we will stick to cosine similarity, while the evaluation of different measures in this context presents an opportunity for future research. Cosine-Similarity measures the difference between vectors by calculating the cosine of the angle between them (Han et al. 2011 p. 42).

The resulting value is bounded in [0,1] since probability vectors only contain elements between 0 and 1 (typically single topics have small probabilities). The cosine-similarity between two vectors **a** and **b** is defined as:

$$\text{Cosine-Similarity(a,b)} = \frac{crossprod(\boldsymbol{a},\boldsymbol{b})}{\sqrt{crossprod(\boldsymbol{a}) * crossprod(\boldsymbol{b})}}$$

A larger cosine-similarity is related to 'more similar' documents and consequently provides a more intuitive scale for similarity, as opposed to a larger angle between the vectors, which is related to less similar documents. If a topic has a high (or low) probability in document A and document B, there might be a relationship between these assignment probabilities. However, the inverted case is unlikely to occur. When comparing the documents in our analysis, we utilize the topic probability vector for the documents, i.e. compare $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_N)$ to $\boldsymbol{\beta} = (\beta_1, \dots, \beta_N)$ using the cosine measure. However, since we are not concerned with the relationship between a single conference call and a single analyst report, but rather the relation between the two content types in general, a second step of aggregation is necessary before the analysis can take place. In order to assess if there is a change in the topic structure of analyst reports following conference calls (C), we select a sample of reports prior to and following each conference call. For both periods we compute the average of the samples' respective topic to document probabilities and denote the average topic probabilities prior to the call as $\boldsymbol{\beta}^-$, while referring to the post-call average as $\boldsymbol{\beta}^+$. If the call has an impact on the topic structure of the reports written after it has taken place, the similarity between the calls topic structure and the reports should increase. If there is no relation between the two content types at all, no stable pattern should emerge.

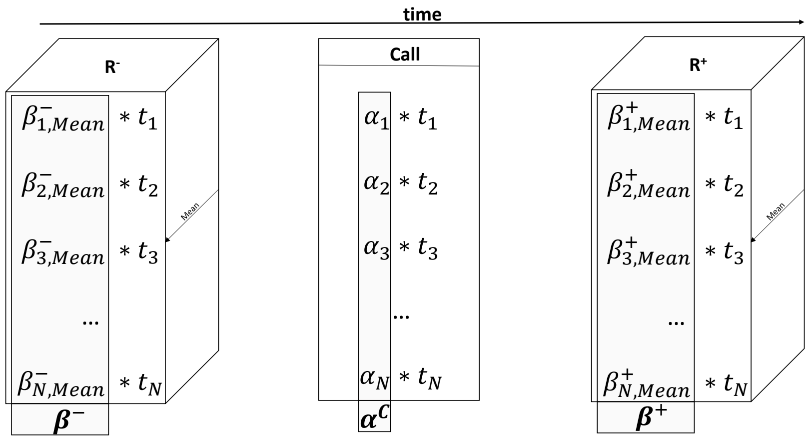

Figure 1. *Illustration of sample selection surrounding a conference call. $\boldsymbol{R}^{+/-}$ refer to the pre-call and post-call report samples. The topic-to-document probabilities of the call are denoted as $\boldsymbol{\alpha}^C$, the average topic-to-document probabilities of the pre- and post-call samples are denoted as $\boldsymbol{\beta}^-$ and $\boldsymbol{\beta}^+$ respectively. Each t_n refers to a topic.*

This leaves us with the task of determining the appropriate sizes of the pre- and post-call report samples. There are, roughly, one hundred reports written between the occurrence of two calls (the number varies), and consequently a pre- and post-call sample of 50 reports each would capture the mid-point between two calls. However there is the concern that if too many reports are selected in the sample, a present relation between the content types might be missed because, in example, only the reports written in the week prior to and after the call may relate to the call. However, if the sample is chosen too small, one cannot be certain that this doesn't miss reports of especially well-informed analysts (pre-call similarity, had a topic early on), or reports written very thoroughly after the call (simply take longer to create). Therefore we conduct the analysis for a large amount of sample sizes and compare the results regarding the mean of pre- and post-call topic similarities to all calls on average. The results presented in the next section exhibit interesting patterns based on this parameter.

5 RESULTS

To give a meaningful impression of the results for the averages described at the end of the last section, we begin the presentation of our results by reporting one example of the data that constitute the average results. Table 2 provides the pre-call ($Cos(\alpha^c, \beta^-$) and post-call ($Cos(\alpha^c, \beta^+)$) averages for a report sample size of 10 reports in each direction from the call.

Call ID	$Cos(\alpha^c, \beta^-)$	$Cos(\alpha^c, \beta^+)$	Change	Call ID	$Cos(\alpha^c, \beta^-)$	$Cos(\alpha^c, \beta^+)$	Change
1	0.079	0.326	+	25	0.003	0.322	+
2	0.074	0.040	-	26	0.045	0.592	+
3	0.040	0.064	+	27	0.022	0.302	+
4	0.047	0.023	-	28	0.020	0.015	-
5	0.037	0.695	+	29	0.020	0.271	+
6	0.026	0.089	+	30	0.338	0.023	-
7	0.246	0.041	-	31	0.016	0.623	+
8	0.045	0.023	-	32	0.100	0.235	+
9	0.332	0.025	-	33	0.027	0.285	+
10	0.042	0.262	+	34	0.033	0.011	-
11	0.115	0.078	-	35	0.294	0.007	-
12	0.043	0.071	+	36	0.054	0.238	+
13	0.025	0.018	-	37	0.057	0.344	+
14	0.138	0.106	-	38	0.044	0.472	+
15	0.046	0.034	-	39	0.025	0.444	+
16	0.045	0.188	+	40	0.075	0.209	+
17	0.017	0.047	+	41	0.125	0.014	-
18	0.033	0.049	+	42	0.017	0.057	+
19	0.033	0.298	+	43	0.073	0.021	-
20	0.007	0.016	+	44	0.020	0.019	-
21	0.022	0.293	+	45	0.507	0.095	-
22	0.008	0.030	+	46	0.025	0.211	+
23	0.004	0.018	+	47	0.031	0.028	-
24	0.005	0.007	+	48	0.070	0.353	+

Table 2. *Mean Cosine-Similarities between each conference call and the mean topic to document probabilities ($\boldsymbol{\beta}^-$and $\boldsymbol{\beta}^+$) for 10 pre- and post-call reports. Positive difference in grey, negative in white. The call IDs are ordered in time, i.e. call 1 is the first call in the sample and call 48 the last.*

As Table 2 shows the pre-call similarity tends to be lower than the post-call similarity for this sample size. The mean pre-call similarity across calls is 7.395% and the mean post-call similarity is 16.737%. The mean-difference is statistically significant on a 99% confidence level. This creates the question how this difference depends on the size of the pre- and post-call report samples. To answer this question, we calculate the mean pre- and post-call similarity for sample sizes between 2 and 300 reports before and after the call.

As shown in Figure 5, the post-call similarity peaks immediately after the call and continues to be larger than the pre-call similarity throughout the chosen report sample sizes. The fact that the difference is significant on a 90% confidence level up until over 100 reports after the call is interesting in itself, as a report sample of this size may very well include the next call. Still, there is no notable peak in similarity for either pre- or post-call similarity, which indicates that topics from one call will typically not be taken up in the next one.

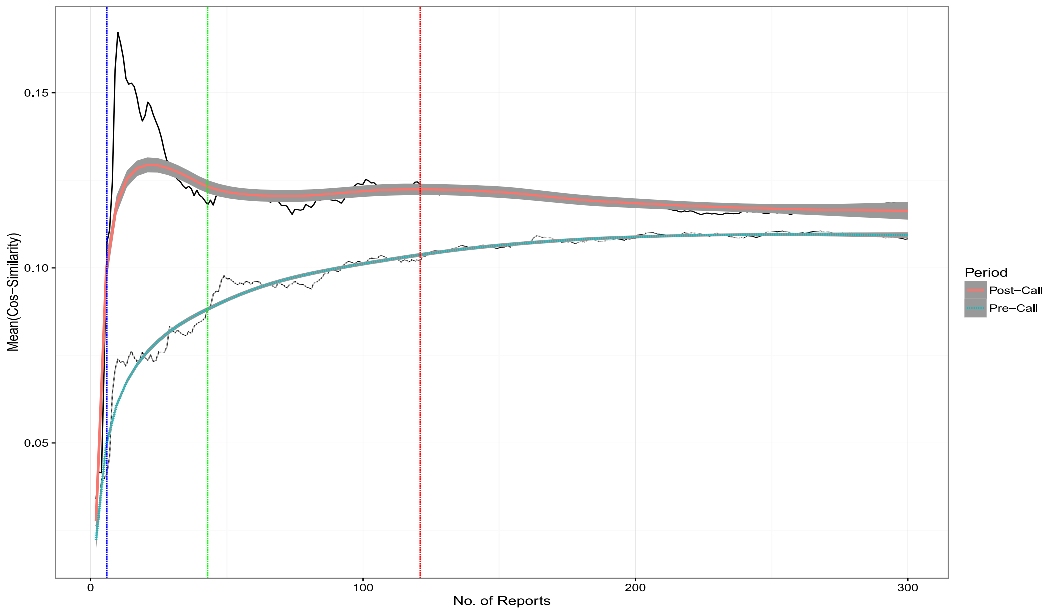

Figure 2. *Average (across all calls) of means (across $\boldsymbol{\beta}^{-}$and $\boldsymbol{\beta}^{+}$) cosine similarities for report samples of varying size n= [2,…,300]. The peak of Pre-Call is located around n=[2,…,20]. The blue vertical line indicates the beginning of statistically significant differences. The green line indicates the end of 95% confidence. The red line the end of 90% confidence.*

More importantly, the peak in post-call similarity is in line with the assumption that analysts are provided with valuable new information during conference calls, which leads to a topic change in post-call reports. However, it is important to keep in mind that we have not tested this against cases were there was an earnings announcement but no conference call. It may be that the shown change in report topics would be similar without the conference event.

5.1 Limitations

There are a number of limitations, which should be kept in mind when considering the results of this research. First, while the sample analyzed is quite large regarding the number of analyst reports and of reasonable size (n=48) regarding the calls included in the analysis, we examine the case of a single firm in this study. While there is a methodological reason for this (a cross-company trained topic model would be unlikely to work well for our analysis), validation of these results in future research using different samples is desirable.

5.2 Future Research

As topic mining is relatively unexplored in IS research compared to other text mining approaches, it does not surprise that more questions remain unanswered regarding the usability of the topics for data mining in general and the implications of topic mining in the domain studied in this research in particular. We identify three possible types of extension of the presented research. First, a similar analysis could be applied to other non-financial content domains. Second, single topics can be explored regarding their individual domain-transfer behaviours. Third, more than two content types could be analyzed simultaneously:

Transfer to other content domains: The approach discussed in this research may be useful in exploring information processing in other application domains. For example, marketing research regarding the adoption of topics from campaigns in social media posts or the analysis of political debates by the examination of topic proliferation in different content types might benefit from a similar analysis. The main requirement for this type of analysis is the existence of historical data that may be used for training the LDA model and that is suitable to be categorized into separate categories.

Introspection of individual topics: Of course, the presented results pose the question which kinds of topics are most likely to be transferred from one content type to the other. LDA does provide the necessary output for this kind of analysis. In example, an inter-author coding of the top words (most probable) in the topics with the highest pre- to post-call volatility could reveal which types of topics are responsible for the observed effects. Both directions of possible topic-transfer are potentially interesting for this kind of analysis. The question which kinds of topics are more likely to be picked up in a call when they are contained in a pre-call report can contribute to further understanding of analyst herding behaviour, while the question which topics are most likely to transfer from calls to future reports could inform analysts which call topics are especially valuable information. This may also help to understand in which situations analysts do not herd, i.e. are especially ‘daring’. Finally, for both types of topics more likely to transfer from one domain to the other, the question of ‘topic value’ may be explored by relating these topics to abnormal stock returns during or after the call events. Another interesting question regarding the behaviour of individual topics is given by the question if topics introduced by the analysts in the Q&A part of the call, which were not present in the presentation part of call, are more likely to have a long lasting impact on the topic structure of reports or an immediate market reaction.

Expansion of current approach to more content types: While the presented analysis reveals interesting patterns of topic-transfer between the content types, it would be daring to assume a causality between the two types of content in a general sense. Financial analysts are, as their name suggests, interpreters of information and not originators of events. Thus, it would be interesting to introduce more content types to this analysis and investigate in which other content types new topics arise before they are picked up by financial analysts or the corporate representatives present during conference calls. There are a number of promising candidates for this expansion of the current analysis, such as news media, social media, corporate filings or regulatory announcements.

6 CONCLUSION

In this research we apply LDA, a topic mining algorithm, to analyst reports and conference calls, in order to investigate financial analysts’ information processing behaviour. To this end, we collect a consecutive sample of reports and calls about IBM covering the period from 2000 to 2015. Keeping in mind the limitations discussed in the previous section, we examine the topic to document assignment probabilities resulting from this model and determine average topic similarities between report samples prior to and after each conference call in the sample, while varying the scope of the pre- and post-call samples. Results indicate that, although individual topics exhibit different behaviour, on average, analyst reports written in a short period after conference call events show a significant topic-uptake from conference call events. This finding is in line with the consideration that analyst conference calls are a valuable source of new information for stock analysts. On the other hand, there is no similar spike in to-call similarity regarding analyst reports released prior to call events, which may be seen as support for the “herding” tendency of analyst opinion, i.e. the tendency to stick to the consensus estimate until new information has reduced the risk of changing ones’ opinion. Future extensions of this study may include the extension of the approach to more companies, which may allow to examine industry differences in topic-transfer between analyst reports and conference calls, content domains, the introspection of the effects of individual call topics, as well as the integration of other measures, such as conference call sentiment and its relation to call topics.

References

Baker, L. D. and A. K. McCallum. (1998). Distributional clustering of words for text classification. In: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR ’98 (pp. 96–103). New York, New York, USA: ACM Press.

Blei, D. M., A. Y. Ng and M. I. Jordan. (2003). Latent dirichlet allocation. The Journal of Machine Learning Research, 3, 993–1022.

Clement, M. B. and S. Y. Tse. (2005). Financial analyst characteristics and herding behavior in forecasting. The Journal of Finance, 60(1), 307–341.

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.

Grün, B. and K. Hornik. (2011). topicmodels : An R Package for Fitting Topic Models. Journal of Statistical Software, 40(13), 1–30.

Han, J., M. Kamber and J. Pei. (2011). Data Mining: Concepts and Techniques, third Edition (The Morgan Kaufmann Series in Data Management Systems). Morgan Kaufmann.

Hong, H., J. Kubik and A. Solomon. (2000). Security Analysts’ Career Concerns and Herding of Earnings Forecasts. The RAND Journal of Economics, 31, 121.

Irani, A. J. and I. Karamanou. (2003). Regulation Fair Disclosure, Analyst Following, and Analyst Forecast Dispersion. Accounting Horizons, 17(1), 15–29.

Lee, M. D., B. Pincombe and M. Welsh. (2005). An Empirical Evaluation of Models of Text Document Similarity. In: XXVII Annual Conference of the Cognitive Science Society 1254–1259.

Sunder, S. V. (2002). Investor Access to Conference Call Disclosures: Impact of Regulation Fair Disclosure on Information Asymmetry. SSRN Electronic Journal.

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, 1–21.