Summer 6-27-2016

HOW TO CONQUER INFORMATION OVERLOAD? SUPPORTING FINANCIAL DECISIONS BY IDENTIFYING RELEVANT CONFERENCE CALL TOPICS

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

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

HOW TO CONQUER INFORMATION OVERLOAD? SUPPORTING FINANCIAL DECISIONS BY IDENTIFYING RELEVANT CONFERENCE CALL TOPICS

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

The ever rising amount of business communications results in a growing amount of qualitative data relevant to many decision situations. This increase in information volume and velocity threatens to overburden decision makers. We provide a structured approach towards this problem using topic-models to reduce information overload by filtering content and by providing context-relevant information to decision makers. Building upon theoretical considerations related to phases of the decision process established by Herbert A. Simon, we implement the proposed approach on the example of a large document collection of stock analyst reports and analyst conference calls using Latent Dirichlet Allocation (a topic model). Thereby, we extract investment-relevant topics from the model and discuss the opportunities for decision support resulting from the chosen approach.

Keywords: Information Overload, Topic Mining, Conference Calls, Latent Dirichlet Allocation.

1 INTRODUCTION

Due to the exponential rise in available unstructured business communications, market participants are faced with an increasingly complex decision problem when new information becomes available to them. Important questions regarding a newly available document are what kind of information is contained therein and what topics should the reader pay the most attention to. In recent years, topic mining models have become readily available and advances in the implementation and processing speed have made it possible to use these models on large document collections. However, determining the topic composition of a document does not in itself provide useful information to a decision maker. We investigate the decision making process on the example of investment decisions on the basis of qualitative (textual) data, which threatens to overburden investors information processing capabilities and may lead to information overload. In order to help investors to arrive at an informed investment decision on the basis of such document topic mixtures one also needs to know which topic is important because it will lead to a particular reaction of the capital market. We outline a procedure for the identification of topics relevant for predictive purposes and showcase this approach using a large corpus of stock analyst communications regarding the firms contained in the Dow Jones Industrial Average (DJIA) index. On the basis of this example, we contribute to the growing research stream of data analytics addressing the problems resulting from information overload. We structure this research as follows. In the section following this introduction we describe our motivation for this research based on the information value of corporate disclosures and the resulting growing amounts of qualitative data, which threatens to overburden decision makers in the context of investment management. On this basis, we describe how the investment decision problem can be incorporated into existing decision-making theory and derive the research questions of this paper based on this background. We build upon the resulting decision-making process to structure our analysis. In the third section of the paper, we introduce the data used throughout the analysis and provide brief introductions to the methodology of the paper, which include topic-mining and the computation of abnormal stock returns. Using this methodology, the fourth section begins with a brief manual 'sanity check' of the chosen approach, before continuing to implement an automated topic selection approach. The presentation of this analysis also includes descriptions of the topics selected by the chosen approach and their impact on abnormal stock returns as estimated by the model. The fifth section of the paper discusses these results regarding the problem of information overload and briefly summarizes the theoretical and practical implications of the analysis. The final section concludes the paper.

2 THEORETICAL BACKGROUND

In this section, we motivate the subject matter of this paper by giving an overview of relevant domain literature regarding the information content of corporate disclosures, of which analyst conference calls are an example often cited especially in accounting literature. Following this introduction to domain literature, we discuss the problems arising from this unstructured data in the context of IS theory. Finally, we introduce Simon's (1977) seminal model of a rational decision-making process providing the theoretical foundation for our research.

2.1 Information Value of Corporate Disclosures

The information value of analyst conference calls beyond the information disclosed in quarterly earnings reports has been studied in a growing stream of literature. Typically, these voluntary augmentations to quarterly earnings announcements consist of a presentation by the management of the company holding the call and a Q&A section during which analysts can ask questions about the announcements or other topics of interest. Scientific interest in the disclosures made during these calls has risen since the mid-1990s when conferences became a popular medium for disclosure. As noted by Li, the understanding of the textual part of corporate disclosure is vital for three reasons (Li 2010): First, the textual portion of disclosures is needed to understand the numeric part of the disclosure. Second, the communication patterns revealed in the unstructured part of disclosures may allow inferences beyond the intended content of the disclosure, such as managerial characteristics. Third, textual content allows insights regarding the managerial point of view, allowing a deeper understanding of firm behaviour. Early work regarding the subject focusses on the question when firms opt to hold calls and suggest that calls are used as a way to increase information quality and as a response to perceived low accounting quality (Feldman 1996; Tasker 1997, 1998; Frankel et al. 1999). More recent work in accounting focusses on the incremental information content of the calls beyond the accompanying press release about earnings figures. Here, the information disclosed during the calls is measured using word list based textual measures (Matsumoto et al. 2011) and the effect of scripted answers to analyst questions is assessed using in-call cosine similarity of textual measures comparing the presentation to the QA section of the call (Lee 2015). As this brief overview of the growing literature regarding analyst call disclosure illustrates, the field is actively seeking ways to extract more information from these calls and regards them as a valuable source of information. However, the unstructured nature of the data and the continuous rise in available business communications contribute to an ever more difficult decision problem for those who need to analyze this textual data as will be described in the next section.

2.2 Big Data and the Risk of Information Overload

As a result of the diffusion of information systems throughout society in general and the economic domain in particular an ever increasing amount of information becomes available to decision makers. While the advent 'Big Data' has created new challenges for information systems as a discipline and practitioners alike, it also has created new opportunities for both (Chen et al. 2012). Also, the challenges it poses, while certainly elevated to a new extreme, are not unprecedented in nature. Indeed, a crucial task of information systems has always been to help the user in the processing of large volumes of information, which otherwise elude human understanding. However, the risk to human understanding resulting from the automated processing of large volumes of information was noted early on (Simon 1976 pp. 285–286), who warned that we may drown in the information produced by the very systems created to further our understanding. This phenomenon of continuously increasing amounts of information sources, which often include duplicate information, has been called an **information overload** and has been a growing concern not only for managerial decision makers but also for the public. The study of the limits to our information processing capabilities has a long history in psychology (Miller 1960) and the potentially dangerous role of information systems regarding this has been studied in numerous publications ever since the internet and electronic communications play an increasing role in our daily lives (Whittaker and Sidner 1996; Shenk 1997; Edmunds and Morris 2000). It also continues to be a field of active research, especially regarding the effects of e-mail related overload (Marulanda-Carter and Jackson 2012; Kushlev and Dunn 2015; Zhang et al. 2015). As the focus of these works show, information overload seems to be especially easy to achieve when textual data is relevant. Another perspective on the information overload problem is given by Shirky, who notes that "It's not information overload. It's filter failure." (Shirky 2011). The subject matter of this research is such a filter mechanism in the context of textual business communications. This leads to the question how such filters may be developed. The next section suggests a structured approach to this question.

2.3 Topic Importance as a Decision Problem

The assessment of newly released business communications regarding their importance for future market developments is, at its heart, a decision problem: Which topics contained in a document inform the reader about the expected future value the company issuing the document? When faced with a complex decision problem, a systematic approach to decision making can help us to divide the problem into several smaller challenges, each of which may be addressed individually. A fundamental model of the decision process was introduced by Simon, who divides this process into three phases (Simon 1977). Since, this model has been augmented and adapted to the decision support context (Sharda et al. 2014). The **intelligence phase** is characterized by two steps. First, the search for the actual problem in the context of organizational goals. As noted by Sharda et al., qualitative (unstructured) data and information overload make this especially difficult (Sharda et al. 2014 p. 75). The growing amount of available information due to the previously outlined developments, results in ever-increasing complexity of the decision problem when trying to identify the relevant contents of business communications resulting in the need for computerized decision support. Second, the search for a well-defined problem statement, which enables decision makers to use their resources effectively by focussing their efforts on the actual problem.

Consequently, the output of this phase is such a well-defined problem statement, which forms the basis for further work in the **design phase**. This second phase again consists of two distinct steps. First, the information required to address the identified problem is gathered. Second, a set of alternatives is developed. As noted by Sharda et al., in the context of decision support, this usually involves the creation of a model in order to reduce complexity and consequently make the problem solvable within the constraints of this model (Sharda et al. 2014 p. 77). Another important aspect of this model-based approach to decision support is the development of what Sharda et al. call a **principal of choice**, a criterion on which to base the answer to the question of what an acceptable solution actually is. Thus, the output of this phase is characterized by two components. A set of possible solutions and a criterion on which choose among them. These components constitute the input of the third and final phase of the decision process, the **choice phase**, in which the alternatives are compared on the basis of the chosen criterion. As the ensuing analysis will show, the distinction between design and choice phase may often be blurred in the case of data analysis, as the process of modelling data is usually iterative or there may be no single optimal choice. Still, the theoretical separation of the phases helps in the conceptualization of solutions as the one presented here. This model of the decision process relates to our research as follows (Figure 1). In the intelligence phase, we determine that decision makers (i.e. investors) face information overload resulting from high volumes of qualitative data. Consequently, during the design phase decision makers need to create and evaluate topic models for filtering non-relevant content from their analysis. Finally, in the choice phase of the decision making process, the most relevant topics regarding a domain appropriate target have to be selected from these models.

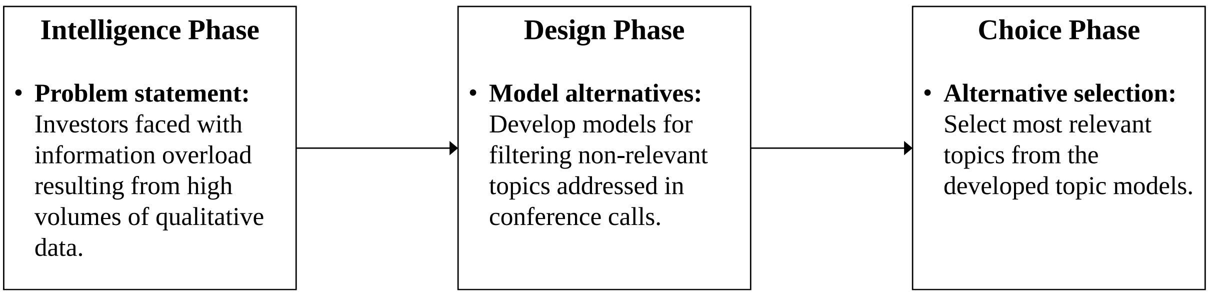

Figure 1: Phases and their outputs based the Decision Process according to (Simon 1977).

Based on these theoretical considerations, we develop the following research questions. The central problem of this decision process is the development of a topic-mining workflow that identifies relevant topics from a large document collection so that information overload can be mitigated. As the course of our analysis will illustrate, three questions need to be answered during the development of this workflow. First, how can 'relevance' of topics be operationalized in the context of our problem domain in a way that is both useful in the decision context and supported by prior research? Or, put concisely:

RQ1: How to define 'relevance' in the context of business communications?

Second, in order to be able to reduce the amount of 'topics' relevant for decision-making, the topics contained in the document collection need to be extracted using some automated approach. During the creation of such a topic model, a number of different decisions need to be made regarding the training-data and scale of the model, which can be expected to have a crucial impact on the effectiveness of the following analysis. These operational questions are addressed by RQ2.

RQ2: How to train topic models in order to obtain relevant and filter out non-relevant topics for predictive purposes in the context of business communications?

Third, once the topic structure of the decision documents relevant within the decision context has been modelled, we need to reduce the complexity of this structure in order to reduce the information overload the decision maker is faced with. Consequently, we require an approach able to automatically select the subset of topics that is relevant according to the criterion developed regarding RQ2:

RQ3: How to identify the 'relevant' subset of topics suitable as information regarding the decision problem?

These research questions correspond with the output of a decision makers' intelligence phase. The problem has been identified on the basis of prior literature and a clear problem statement with solvable steps toward a solution has been the result. Following these theoretical considerations, we provide an approach to solving the problem of document topic importance on the example of analyst to company communications. The following section provides an introduction to the data and methodologies used in our analysis.

3 DATA AND METHODS

Consequently, the design phase of the decision making process begins with its first step: The gathering of information relevant to a possible solution. Information is defined twofold in this context. First, the data necessary for a solution needs to be collected. Second, suitable methodology needs to be identified. We collect a total of 117,398 documents regarding the 30 companies contained in the Dow Jones Industrial Average (DJIA) covering the years 2000 through May 2015. The document collection consists of two kinds of documents. The smaller portion of the collection is given by analyst conference call transcripts. Typically, these calls are held when quarterly and annual earnings are announced, resulting in five regular calls per year and company. In addition, calls may be scheduled for significant developments. This results in a total of 2,643 call transcripts in our sample. These calls are used as the main subject of our topic mining analysis. The remaining 114,755 documents are analyst reports regarding the companies. Typically, such reports include an opinion on the current financial state of a company, as well as a buy/hold/sell type recommendation regarding the firm's stock. However, reports may also cover any other subject of interest to the customers of analysts. It is due to this variety in subjects and the relatively large number of available documents we choose to include these reports as training data for the topic model. The same reason leads to the selection of the 30 DJIA constituents as our company sample. The size of the companies included in this index guarantees comprehensive analyst coverage. Both document types are obtained from Thomson Reuters Advanced Analytics. Both document types are pre-processed by removing *stop words*, such as *it, he, and* or *to*. We also remove a number of recurring patterns from the documents, such as the legal disclaimer appended to each call transcript as these would otherwise dominate the topic structure. The texts are consequently transformed into a word vector representation preserving the word order of the documents before processing them using the method outlined in the following section. In the following textual analysis, we use a number of different statistical software packages (McKinney n.d.; McCallum 2002; Wickham 2009; Friedman et al. 2010; R Core Team 2015). In addition to this textual data, and in order to determine the 'relevance' in the context of business communications (RQ1), we collect the corresponding set of stock price data for the firms included in the DJIA from Thomson Reuters DataStream.

3.1 Topic Mining using Latent Dirichlet Allocation:

In this section we introduce the chosen topic mining algorithm before continuing to the estimation procedure for the return model used to identify the subset of topics in the next subsection, which is useful for predictive purposes. Latent Dirichlet Allocation (Blei et al. 2003) is a topic mining algorithm that considers a document as the result of a generative process of topic mixtures. In turn, a topic is viewed as a mixture of words. Consequently, the model outputs assignment probabilities for all words to all topics ($\mathbf{M}_T$) and for all topics to all documents ($\mathbf{M}_D$) in the corpus. More formally, and following the notation of (Blei et al. 2003), for a corpus **D,** defined as a vector of documents **w**, each in turn consisting of **N** individual words w_n, the topic mixtures Θ are computed as follows:

1.	Choose **N** ~ Poisson(ξ)	
2.	Choose Θ ~ Dirichlet(α)	
3.	For each of the **N** words w_n:	
	I:	Choose topic z_n ~ Multinomial(Θ)
	II:	Choose a word w_n from $p(w_n \vert z_n, \beta)$, i.e. the multinomial conditional probability of the word conditioned on the topic z_n.

Table 1: *Mock-algorithmic description of topic mining via Latent Dirichlet Allocation* (Blei et al. 2003).

As noted, this results in two matrices. Here we refer to these as $\mathbf{M}_D$, the topic to document probabilities, and $\mathbf{M}_T$, the word to topic probabilities. For our purposes, the topic to document matrix $\mathbf{M}_D$ is the focus of interest. Still, the word to topic matrix can be used to inspect the individual topics regarding their meaning if such introspection is desired. A number of decisions regarding this model have to be considered regarding our predictive goals. First, the number of topics the model generates needs to be chosen a priori. If this number is chosen too small, the resulting topics lack granularity. For example, if only 10 topics were trained for our sample of 30 companies, topics would likely only decide between different industries included in the sample. If it is chosen too large, the topics will not be divisive. As we expect only a portion of the resulting topics to be relevant (in terms of stock market reaction, which will be introduced in the following section), the chosen number of topics should also be taking this into account. A final consideration is the relation between sample size and topic count. As we include a large number of documents in the training sample of the topic model, we are not constrained by this. However, larger topic counts increase the training time needed to create the model. Consequently, we train a model including 100 topics, i.e. a little over three topics per company in the sample. Furthermore, which documents the model is trained on needs to be decided. As we are interested in the stock market reaction to analyst conference calls, we need topics that enable us to capture the contents of these documents. However, these calls are comparatively rare. Our sample of 2,643 call transcripts regarding 30 companies may simply be too small to train the model. Indeed, we trained a model using only calls as input data for comparison. This model performed much worse than the chosen approach. This addresses RQ2, i.e. how to train a topic model that contains relevant topic information.

3.2 Assessing Stock Market Reactions

In order to assess the suitability of the topics generated by LDA for predicting price adjustments at stock markets, which we use as a definition of "relevance" in the context of business communication, we need a suitable measure for such reactions to analyst conference calls. This addresses RQ1, i.e. the development of a decision-relevant measure for topic 'relevance'. We use event study methodology (MacKinlay 1997) to measure this reaction around each call date. In short, an event study measures the abnormal component of stock returns surrounding an event (i.e. the conference call) by establishing a baseline for these returns using historical data prior to the event (estimation window) and comparing this *normal period* to an *event period* (event window, including the call date). The underlying assumption of this method is that market participants will quickly react to new information by adjusting their evaluation of a company. The aggregate change of opinion about the company should in turn be reflected immediately in the price of the company's shares. The efficient market hypothesis is commonly used as a theoretical foundation for this process (Fama et al. 1969; Malkiel and Fama 1970). The abnormal return $AR_{i,t}$ for asset *i* at time *t* is defined as follows.

$$AR_{i,t} = R_{i,t} - E(R_{i,t} \vert X_t)$$

where $E(R_{i,t} \vert X_t)$ represents the (expected) normal return with X_t, representing the conditional information for a normal performance model and $R_{i,t}$ represents the actually observed stock return

(Campbell et al. 1997). These abnormal returns are consequently cumulated over the event window *[t₁, …, T]*, yielding the cumulative abnormal return $CAR_{i,t}$.

$$CAR_{i,t} = \sum_{t=1}^{T} AR_{i,t}$$

We estimate the normal returns $E(R_{i,t}|X_t)$ using the market model approach (MacKinlay 1997), which assumes a constant linear relationship between asset individual returns and X_t represented by a market return (S&P500 in our case) and estimate the relationship via OLS regression.

$$R_{i,t} = \alpha_i + \beta_i R_M + \epsilon_{i,t} \text{ with } E(\epsilon_{i,t}) = 0 \text{ and } Var(\epsilon_{i,t}) = \sigma^2_{\epsilon,i}$$

These $CAR_{i,t}$ will be used to assess the impact of each topic on the market reaction surrounding each call event. Those topics that lead to a significant price reaction are then defined as relevant on the context of business communication. To provide an overview of how the different calls impact the stock price on average (*CAAR*), we group the call events by call sentiment quantiles and define the central two quantiles as neutral and the uppermost und lowest quantile as positive and negative respectively. Figure 2 shows the result of this grouping. Call sentiment is calculated using the McDonald dictionary (Loughran and McDonald 2011). We calculate sentiment using the following dictionary-based measure *Mood* for document *D*, in which *Pos* and *Neg* refer to the positive and negative categories of the dictionary. In a later stage of the analysis we split the analyst calls into several parts, Q&A section and presentation, for each of which we again compute separate *Mood* measures based on whether a contribution is given by an analyst or an employee of the company holding the call (Table 4).

$$Mood_D = \frac{Pos_D - Neg_D}{Pos_D + Neg_D}$$

As illustrated, the CAARs start to show movement before the actual call events. Thus, we opt to use CARs (not averaged) starting one day before the call for further analysis. In addition, the effect of the call events seems to fade away after approximately 10 days. Consequently, we focus on event windows ranging from *[-1, -1]* to *[-1, 10]* days surrounding the calls resulting in 12 possible CAR lengths.

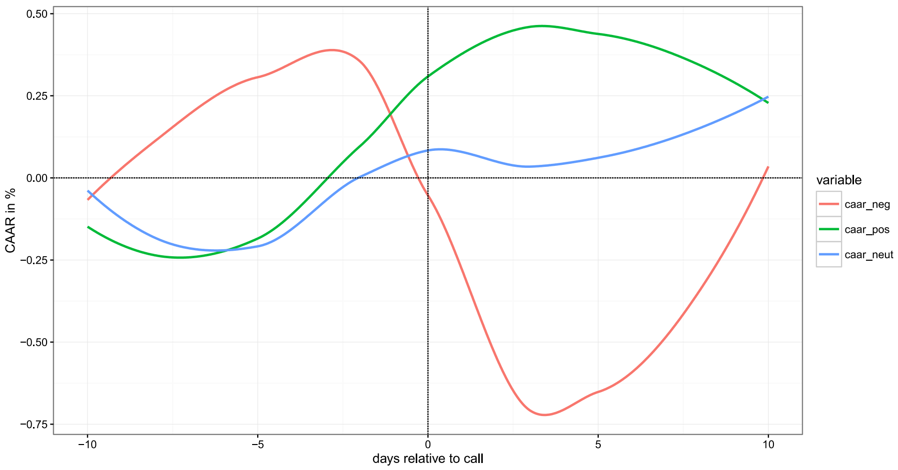

Figure 2: *Cumulative average abnormal returns (CAAR) surrounding analyst conference calls grouped by Mood quantiles. The positive category contains the 25% most positive calls, the negative category the 25% most negative calls and the neutral category the remaining 50% between those two categories.*

4 ANALYSIS

After the necessary data and methodology for our analysis has been introduced, we continue with the second step of the **design phase**, i.e. the development of alternatives. Technically, these alternatives are already developed as we have trained the topic model containing all 100 topics. However, in order to move on to the choice phase and the evaluation of the alternatives, we first need to answer two practical questions. First, which topics (if any) are suitable predictors in models explaining the abnormal returns surrounding the call events? Second, which event window (of abnormal returns) is suitable for which topics? As mentioned, we trained 100 topics in our model, each of which needs to be assessed for suitability regarding the 12 possible CAR lengths. In order to gain a first impression before solving this problem with a computational approach, we opt to manually investigate if the topics are suitable predictors at all. Figure 3 shows the p-values of a smaller 20-topic model using the different CAR-lengths as dependent variables. The lines shown in the figure are interpolations making the graph more readable. As shown, even in a naïve model containing all topics, several topics are indeed suitable as predictors for call CARs (i.e. we consider them as relevant) for a number of different CAR periods (below dashed red line). Another important observation is given by the fact that most topics are not suitable for our purposes (above dashed red line). This is in line with intuition, after all most topics discussed in analyst conference calls should not be surprising to the market and consequently do not reflect new information which might trigger a market response. It is also in line with the goal of the analysis, i.e. the reduction of the amount of relevant content in order to mitigate information overload.

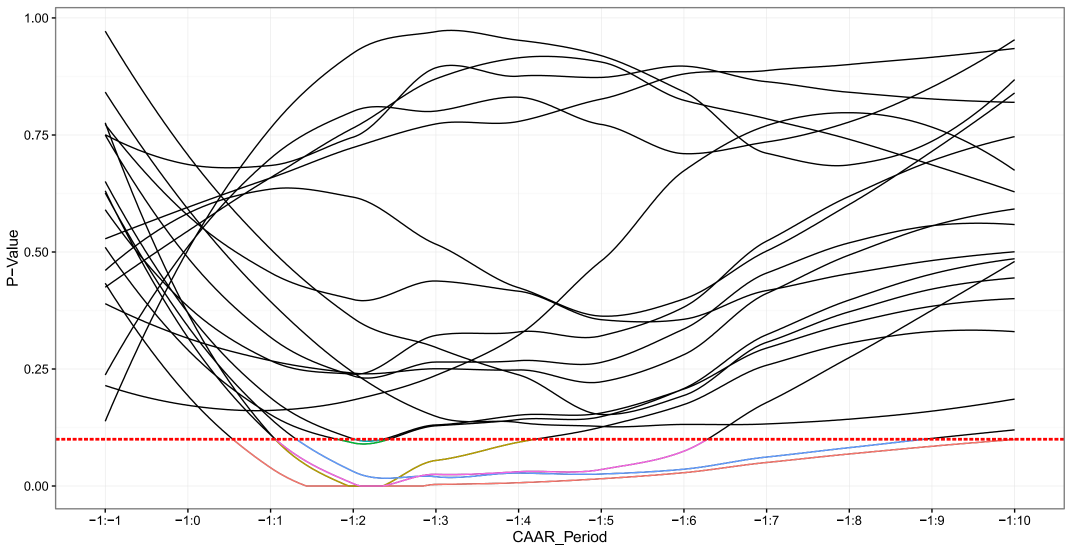

Figure 3: Graphical representation of p-value traces for a topic model across different CAAR measurement periods surrounding analyst conference calls, ranging from [-1:-1] to [-1:10]. For example, [-1:2] refers to cumulative abnormal returns measured starting one day before a conference call to until two days following the call, i.e. CAR(-1,1). The dashed red line indicates Pr(>|t|) ≤ 10%, i.e. a common minimal requirement for significance. Each other line represents one topic.

However, this observational approach is neither statistically valid nor convenient for larger amounts of topics. Therefore, we opt to select topics based on an automated approach, which corresponds with the first part of the **choice phase** of the decision process, i.e. the evaluation of alternatives. We use the least absolute shrinkage and selection operator (LASSO) approach (Tibshirani 1996) to variable selection, as opposed to ridge regression, because of the assumption that only a small number of topics should be relevant, i.e. they should have an impact on abnormal returns. The LASSO approach has two advantages regarding our topic-selection problem. First, it selects a small amount of estimators by penalizing the absolute value of coefficients and driving them towards zero, thus effectively eliminating them from the model. Second, this also results in the exclusion of highly correlated explanatory variables, which may otherwise be a problem for similar topics (Friedman et al. 2010). The LASSO provides parameter estimates as solutions to a minimization dependant on a parameter λ, which serves as a complexity parameter for the model. The λ parameter controls the strength of the penalty towards higher model complexity, i.e. including more covariates. Consequently, by varying this parameter smaller subsets of the 100 topics are selected into the final model for each CAR length. In order to improve the reliability of the model selection process we perform 100-fold cross-validation. This splits the data into random training and test sets 100 times and cross validates the results of these random splits. Figure 4 shows the development of the mean squared error on the example of models for CAR(-1,0) and CAR(-1,4). The lowest part (first vertical line) of each curve indicates the λ value minimizing the cross-validated mean squared error and consequently the value λ_{MIN} chosen for each CAR length in our analysis. This results in one model per CAR length. Alternatively, the second vertical line marks the point where the error is one std. error above the minimum and can serve as an alternate, more relaxed, criterion. For the sake of simplicity, we focus on the former value.

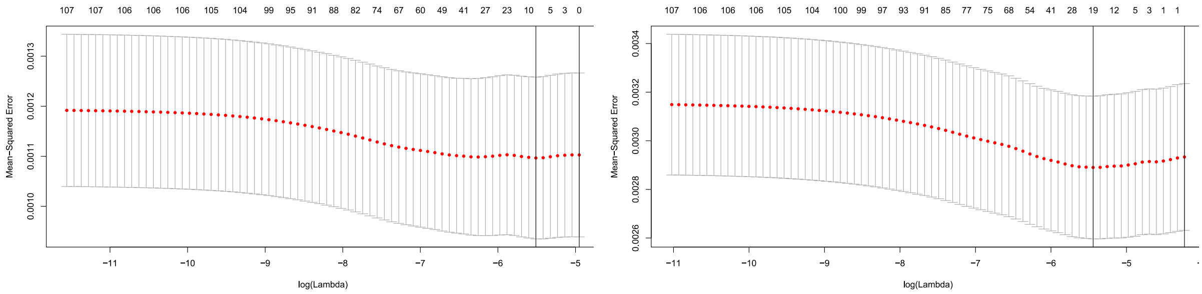

Figure 4: *LASSO procedure traces of the mean squared errors depending on log(Lambda) for models CAR(-1:0) (left) and CAR(-1:4) right. In each plot the first vertical line marks the minimal mean squared error, while the second line marks one standard deviation. Grey error bars indicate the upper and lower end of the curves standard deviation.*

We perform this procedure for each CAR length and consequently determine the set of topics included in the respective chosen models. This addresses RQ3, i.e. how to select decision-relevant topics from the trained topic model using the chosen relevance-criterion.

Table 2 reports the resulting counts of topics included in each CAR period. As shown, there is a peak in topic relevance in CAR periods ranging from CAR(-1,2) until CAR(-1,4). It is important to remember that due to the random nature of cross validation some variation is expected when performing the analysis. However, while there is some variation in the exact position of the peak, results are consistent across repeated cross validation runs as illustrated by the two alternate cross validation runs (No. Topics_2 and No. Topics_3) included in the table. Also, the topic-groups included in other cross validation runs are similar to those included in the one chosen for further analysis.

CAR Period	**-1:-1**	**-1:0**	**-1:1**	**-1:2**	**-1:3**	**-1:4**	**-1:5**	**-1:6**	**-1:7**	**-1:8**	**-1:9**	**-1:10**
No. $\text{Topics}_{\text{used}}$	5	2	0	9	12	15	0	0	0	1	1	0
No. Topics_2	5	2	0	9	12	15	0	0	0	0	1	0
No. Topics_3	2	2	13	9	12	15	0	0	0	1	1	0

Table 2: *Number of topics included in the models chosen by LASSO depending on the CAR aggregation period. The first No. Topics row represents the models used for further analysis. The two following rows show results for alternate cross validation runs.*

After choosing the topics suitable for inclusion in models for the different possible CAR-lengths in this manner two further questions remain in order to arrive at an informed decision about topic-relevance (RQ2-3): First, what are the topics actually included in the models concerned with? Second, how do the individual topics influence the abnormal returns? To answer these questions we select the CAR(-1,2) to

CAR(-1,4) period as a working example for further analysis and report the top words of the topics included in this model. The selection of these models is based on the fact that these CAR lengths consistently include more topics in the chosen models than all other lengths and are consequently more interesting regarding the research question of this paper. Actually, each topic corresponds to a distribution of all words included in the document collection, which can make the top words appear somewhat similar for different topics. The actual information content of a topic is given by the order of importance it gives across all words of the document collection. Nonetheless, curiosity dictates that some "title" should be given to a topic for analysis. We provide such a classification, which we created by manually and independently coding the top words to a topic name before merging the two into the final names reported (Table 3). As any manual treatment of the variables may be highly problematic in a predictive setting, it is important to remember that these "topic titles" are merely a label making the interpretation of the model easier and play no part in its creation.

No.	Topic Title	Corpus	Top 20 Words
8	Pharma, capital market	0.15%	pharmaceuticals pharmaceutical limited laboratories pharma private stock company corporation plc exchange code pvt life number therapeutics industries usaprivate corp group
9	Barclays Rating	0.45%	barclays capital bank rating research investment plc price coverage andor sector banking financial publication stock universe target equity affiliate report
15	Executives, Strategy	0.60%	company financial president senior executive management board vice key analysis report strategic products code corporation services source officer swot chief
29	Risk Rating	0.22%	del risk insight research disclosure ing gruppo pag rating dei nel con axia report medium usd che bias securities buy
31	Technology, Intel	1.00%	intel intc intels market amd technology processor gross corporation ghz revenue semiconductor server corp products processors demand inventory desktop devices
39	Jaffray Estimates	0.17%	jaffray piper report research company analyst securities investors investment price subject davenport time sell target income information analysts ratings buy
42	Energy, Exxon	0.79%	oil production gas chevron exxon mobil exxonmobil cvx corporation xom natural upstream refining earnings prices exploration downstream project integrated projects
47	Telco, Financial Ratios	0.82%	verizon wireless att communications wireline data adds revenue net ebitda growth services arpu fios postpaid iphone service telecom customers broadband
57	Pharma. R&D	0.35%	pharmaceuticals development research corporation agreement license technology collaboration company pharmaceutical deals deal corp plc announced marketing commercialization life products pfizer
58	Argus Estimates	0.12%	argus report research information growth shares billion analysis financial rating page investment eps dividend company estimate share market price expected
64	Legal Disputes	1.79%	court company patent million case filed litigation settlement agreement claims related approximately march federal cases decision june legal costs state
65	McDonalds Food	0.66%	mcdonalds sales mcd restaurants europe restaurant comps company food menu corporation operating day store included year corp samestore costs comp
68	Quaterly Results	1.03%	quarter year business financial call analyst growth good conference continue results strong question percent earnings weve expect part performance time
72	Aerospace	0.80%	boeing aircraft defense commercial company systems aerospace deliveries production program space orders military boeings airplanes delivery programs air cash airbus
80	Pharma. Marketing	0.71%	johnson sales jnj products medical care market growth consumer pharmaceutical product devices drug international company patients risperdal diagnostics jnjs stent
86	Pharma. Technical	0.43%	cancer disease phase products trial infections life treatment science analytics product report pfizer pain published pipeline disorders system indication (removed one nonsensical token)
95	Risk, Financial	0.12%	risk rating ratings credit share price buy line financial rapid current health term high red blue sell medium market grade
98	Macquarie Financial	0.19%	macquarie research capital securities york return usa tel toronto financial stock distributed issued expected canada exchange limited price markets investment

Table 3: Descriptions of topics included in the CAR(-1,2) to CAR(-1,4) models chosen by the LASSO variable selection. The topic title column shows a manual coding of the top words contained in each topic and serves as a legend the model itself. The Corpus column shows the percentage of the corpus, including both analyst reports and conference call transcripts, each topic makes up according to the LDA model.

We add the %-Corpus column to the table as additional information about the contribution of these topics to the corpus in general. As we trained a 100 topic model, any topic above 1% provides an above-average proportion of the corpus. In example, the topics chosen for the CAR(-1,3) model make up a total 6.88% of the corpus. In other words, the vast majority of the topic mixture of the corpus does not contribute explanatory power to the selected models. Considering the huge amounts of text and the fact that predicting future firm performance, for which abnormal returns surrounding the call events are a proxy, is the point of both analyst reports and the analyst conference calls this large amount of "unused" text is a problem onto itself. However, a manual analysis of the same textual content may reveal information beyond the models scope or even a non-overlapping set of information. As shown, a number of different financial topics are included in the topic set chosen by variable selection. Of course, this is unsurprising as financial results are a key indicator for firm performance and quarterly results are the main topic of the analyst conference call transcripts. When interpreting the models suggested by LASSO it is important to note that the models selected by this approach may not be interpreted using ordinary test-statistics. Thus, we do not report any *p*-values or similar evaluations of the parameters. In a recent article the authors of the original LASSO publication (Tibshirani 1996) and the implementation used in this research (Friedman et al. 2010) have proposed a valid test statistic for the procedure (Lockhart et al. 2014).

Target	**CAR(-1,2)**	**CAR(-1,3)**	**CAR(-1,4)**
Intercept	0.037741	0.0572	0.05825
QA Analyst Mood	0.01578	0.01785	0.01589
QA Corporate Mood	0.00709	0.00925	0.0111
Presentation Corporate Mood	0.0062	0.00726	0.00746
Firm Size	-0.0073	-0.0108	-0.011
Topics:			
8: Pharma, capital market	-	-	-0.3424
9: Barclays Rating	-	0.00631	-
15: Executives Strategy	-	0.14434	0.16172
29: Risk Rating	0.12192	-	-
31: Technology, Intel	-0.8192	-1.2912	-1.0714
39: Jaffray Estimates	-	0.54361	1.18669
42: Energy, Exxon	-	-	1.66477
47: Telco, Financial Ratios	-	1.55913	2.13841
57: Pharma. R&D	-	-	-0.7018
58: Argus Estimates	-1.3107	-0.3391	-0.1467
64: Legal Disputes	-0.054	-0.0878	-0.0998
65: McDonalds Food	0.15006	-	-
68: Quaterly Results	-	-0.1052	-0.0913
72: Aerospace	-0.0538	-1.1264	-0.4132
80: Pharma. Marketing	-	-	-0.0958
86: Pharma. Technical	0.63432	0.48977	0.24329
95: Risk, Financial	-0.3094	-0.623	-0.6919
98: Macquarie Financial	0.32437	0.60405	0.75057

Table 4: Effect sizes for the CAR(-1,2) to CAR(-1,4) models selected by the LASSO.

However, as this test statistic is not yet available in the used implementation and the only available implementation is considered experimental by the authors of the paper about the test-statistic, we refrain from using it here. It should be kept in mind that this does not mean that the models produced by LASSO do not include significant covariates. Another obvious idea would be to simply use the selected variables to estimate a model independently of the selection procedure, however, this is invalid because of the same statistical considerations outlined before. Of course, these authors checked if this produces significant results, and it does even when only using a random sample of 50% of our observations, but as these test statistics may be unreliable we do not report them here in the interest of rigor. Once the mentioned test-statistic becomes widely available in implementations this problem will likely be solved. Keeping this in mind, Table 4 reports the effect sizes for the topics included in the chosen models for CAR(-1,2) to CAR(-1,4). *As illustrated, the topics selected by the LASSO are stable across the different CAR-aggregation periods, which is to be expected if the procedure works as intended because of the* *strong correlation between the different CAR-lengths (remember Figure 2). Furthermore, the estimated topic-coefficients are stable regarding their sign and show monotonous developments across the CAR-lengths. The* Mood *variables and* Firm Size *are included as controls for comparison with the topic variables. As the table shows, the estimated coefficients of these controls are neglicable compared to the call topics. This is an interesting result in itself as such descriptive variables are commonly included in models explaining abnormal returns. Whether this holds true for other controls is an interesting question for future research. This comparison aside, the effect sizes of the call topics are both 'sane', i.e. do not contradict rational expectation. In example, the* Legal Disputes (64) *and* Risk Financial (95) *topics lead to a negative market reaction, while* Pharma Technical (86)*, i.e. a factual topic related to operations, shows a postive effect. Also, the amount of topics included in the models is managable with at most 15 topics per model, which equals 15% of the topics in the topic model or an 85% reduction in topic complexity. Consequently, we consider the topic selection approach to be suitable for complexity reduction.* This leaves us with the final step of the **choice phase**. As is so often the case with statistical considerations, it is difficult to give a general answer to what model is 'best'. However, as the chosen covariate-selection approach conveniently identifies one explanatory set of topics per CAR target period, the choice is readily provided by this methodology. Also, it would be inadvisable to use a single CAR period as the 'solution' to the topic-relevance question because the development of the topic importance across the three models is a crucial part of the information generated by the process as, given the investment context of the analysis, changes in expected stock returns between the different periods may be decision relevant. Consequently, the appropriate output of the choice phase is given by Table 4 in the investment related decision context, in which the price change between any two days may be decision relevant.

5 DISCUSSION

The goal of our analysis was to describe the development of a topic mining and selection approach, which helps decision makers reduce their information overload by providing them with a subset of topics contained in a document collection, which is relevant according to a domain-specific criterion. We demonstrated an LDA-based analysis and optimized the topic selection regarding abnormal returns as an investment relevant criterion, which provides the answer to RQ1, i.e. the definition of relevance in the context of business communications. Of course, stock returns are only one possible solution to this question. In other non-investment contexts, this criterion would have to be changed. In example, in a marketing context, product sales or ad-impressions could serve as an appropriate alternative. The important characteristic of the criterion regarding the chosen selection procedure is 'quantifiability', i.e. the feasibility to use the criterion as the object of a statistical optimization procedure. This could also include categorical or binary measures. As the resulting models include a total number of 18 topics, down from 100 total topics in the model, the desired complexity reduction is achieved by the chosen approach. Perhaps as important than the answer to the question which topics are relevant in the decision context is which topics are not relevant in the decision making process.

We do not provide topic descriptions of the remaining 82 topics not included in the model for reasons of brevity, but a decision maker confronted with potential information overload could analyze these in order to gain information about which news to ignore. Of course, not all of these 88 topics may be interesting regarding this goal, but as mental capacity is a limited key resource in decision making, any potential complexity reduction is worthwhile. This leads to the reflection of our results on the broader context of information overload. Any model-based approach to complexity reduction sacrifices some information in the interest of simplification. In example, the chosen topic model considers topics as nothing more than statistical co-occurrences of words in a document collection and cannot provide insights about the relation of these topics to the broader economic context of the described investment problem or the change in topic importance over time. Consequently, while this topic-model based approach can certainly help to reduce information overload during the decision-making process, the corresponding loss of information has to be considered carefully within the confines of the domain-specific decision problem. Within the financial domain, the operationalization of topic relevance is comparatively easy. The goal of investments (profit) is clearly measurable and relatively well-understood. Even slight additions to this goal, such as the incorporation of moral considerations into the investment decision, prohibit the chosen approach to the problem. Still, when the goal is as clear as in the analysed case, the presented approach can provide an important contribution given by the reduction of decision-complexity the model achieves.

5.1 Implications

We decide between theoretical and practical implications of this research. The practical implications of the research are given by the identification of decision-relevant topics and the resulting reduction of decision complexity. Another possible application of the chosen approach is given by the possible automation of the decision process. As no step of the chosen approach requires manual intervention, this could easily be achieved. However, such automation would require extensive back-testing and should incorporate other non-topic based approaches to stock return estimation. The theoretical contribution of this research is given by its addition to the growing stream of literature regarding the risk of information overload and the presented mitigation approach to such problems. While the implementation in this paper approaches the problem from a finance-specific point of view, other domains struggling with potential information overload may benefit the suggested approach, too. Applications in marketing (sales prediction, campaign impact) or e-commerce (review usefulness) offer interesting avenues of future research.

5.2 Limitations

The results of this research have to be considered keeping in mind the following theoretical and practical limitations. First, the chosen model assumes a constant per-topic impact over time. However, market reactions to a given topic may change over time. Similarly, the chosen stock return assumes a linear and time constant relationship between the returns of the benchmark index (SP_{500}) and individual stocks. While the DJIA company sample, according to the index-definition, provides a representation of the US-economy, a larger sample may improve results. As the document collection, especially in the case of analyst reports needed as training data, is time intensive, this is not easily mitigated. Finally, the chosen covariate selection approach prohibits the computation of test-statistics for the coefficient estimates in the resulting models, which, as noted, may be mitigated by future implementations of existing statistical research.

6 CONCLUSION

The goal of this research was to investigate the possibilities topic models offer in the context of the ever growing amount of qualitative data available to decision makers. As all models, topic models reduce decision complexity by imposing an artificial structure upon reality. We show how this complexity-reduction can be used to mitigate the risk of information overload on the example of business communications between stock analysts and corporations regarding the announcement of quarterly results. We use the decision-making model proposed by Simon to structure the decision problem into separate and therefore more approachable steps, which we address consecutively during our analysis (Simon 1977). We provide pointers on how to train topic models containing relevant topics and use stock returns as a relevance criterion from the financial domain, on which we exemplify how to select relevant topics from the topic model. As a result of this selection process we show how decision makers can be supported by reducing decision complexity and by providing per-topic market reactions to corporate disclosures. This leads to several interesting questions for future research. First, the chosen example from the financial domain can be explored further regarding the relation between existing models of market reactions to new information and market efficiency. Second, the broader issue of information overload can be explored further by combining the chosen approach to complexity reduction with other information filters.

References

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

Campbell, J. Y., A. W.-C. Lo and A. C. et al. MacKinlay. (1997). The econometrics of financial markets (Vol. 2). Princeton University Press Princeton, NJ.

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

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.

Fama, E. F., L. Fisher, M. C. Jensen and R. Roll. (1969). The adjustment of stock prices to new information. International Economic Review, 10(1), 1–21.

Feldman, B. (1996). It's Your Call. Investor Relations, December, 35–37.

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.

Friedman, J., T. Hastie and R. Tibshirani. (2010). Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, 33(1), 1.

Kushlev, K. and E. W. Dunn. (2015). Checking email less frequently reduces stress. Computers in Human Behavior, 43, 220–228.

Lee, J. (2015). Can Investors Detect Managers' Lack of Spontaneity? Adherence to Predetermined Scripts during Earnings Conference Calls. The Accounting Review, 91(1), 229–250.

Li, F. (2010). Textual analysis of corporate disclosures: A survey of the literature. Journal of Accounting Literature, 29, 143.

Lockhart, R., J. Taylor, R. J. Tibshirani and R. Tibshirani. (2014). A significance test for the lasso. Annals of Statistics, 42(2), 413.

Loughran, T. and B. McDonald. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35–65.

MacKinlay, A. C. (1997). Event studies in economics and finance. Journal of Economic Literature, 35(1), 13–39.

Malkiel, B. G. and E. F. Fama. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), 383–417.

Marulanda-Carter, L. and T. W. Jackson. (2012). Effects of e-mail addiction and interruptions on employees. Journal of Systems and Information Technology, 14(1), 82–94.

Matsumoto, D., M. Pronk and E. Roelofsen. (2011). What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. The Accounting Review, 86(4), 1383–1414.

McCallum, A. K. (2002). MALLET: A Machine Learning for Language Toolkit.

McKinney, W. (n.d.). pandas: a Foundational Python Library for Data Analysis and Statistics.

Miller, J. G. (1960). Information input overload and psychopathology. American Journal of Psychiatry, 116(8), 695–704.

R Core Team. (2015). R: A Language and Environment for Statistical Computing. Vienna, Austria.

Sharda, R., D. Delen, E. Turban, J. Aronson and T. P. Liang. (2014). Businesss Intelligence and Analytics: Systems for Decision Support-(Required) (10th ed.). Essex: Prentice Hall.

Shenk, D. (1997). Data Smog: Surviving the Info Glut. Technology Review, 100(4), 18–26.

Shirky, C. (2011). It's Not Information Overload. It's Filter Failure. New York.

Simon, H. A. (1976). Administrative behavior (3. ed.). New York: The Free Press.

Simon, H. A. (1977). The New Science of Management Decision. New York: Prentice Hall PTR.

Tasker, S. C. (1997). Voluntary disclosure as a response to low accounting quality: Evidence from quarterly conference call usage. Available at SSRN 2949.

Tasker, S. C. (1998). Bridging the information gap: Quarterly conference calls as a medium for voluntary disclosure. Review of Accounting Studies, 3(1-2), 137–167.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical

Society. Series B (Methodological), 267–288.
Whittaker, S. and C. Sidner. (1996). Email overload: exploring personal information management of email. In: Proceedings of the SIGCHI conference on Human factors in computing systems 276–283.
Wickham, H. (2009). ggplot2: elegant graphics for data analysis. Springer Publishing Company, Incorporated.
Zhang, A. X., M. S. Ackerman and D. R. Karger. (2015). Mailing Lists: Why Are They Still Here, What's Wrong With Them, and How Can We Fix Them? In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems,4009–4018.