Crowd Analysts vs. Institutional Analysts – A Comparative Study on Content and Opinion

Steffen Bankamp
University of Goettingen, Chair of Electronic Finance and Digital Markets, Goettingen, Germany

Nicole Neuss
University of Goettingen, Chair of Electronic Finance and Digital Markets, Goettingen, Germany

Jan Muntermann
University of Goettingen, Chair of Electronic Finance and Digital Markets, Goettingen, Germany

Crowd Analysts vs. Institutional Analysts – A Comparative Study on Content and Opinion

Steffen Bankamp[1], Nicole Neuss[1], Jan Muntermann[1]

[1] University of Goettingen, Chair of Electronic Finance and Digital Markets
Goettingen, Germany
`steffen.bankamp@uni-goettingen.de`
`nicole.neuss@uni-goettingen.de`
`muntermann@wiwi.uni-goettingen.de`

Abstract. The ongoing digital transformation shapes the world of information discovery and dissemination for investment decisions. Social investment platforms offer the possibility for non-professionals to publish financial analyst reports on company development and earnings forecast and give investment recommendations similar to those provided by traditional sell-side analysts. This phenomenon of "crowd analyst reports" has been found to provide an adequate alternative for non-professional investors. In this study, we examine the informational value of these crowd analyst reports regarding their timeliness in publishing and their originality as for content and opinion. Our findings suggest that crowd analysts strongly rely on previously published institutional reports. Therefore, crowd analysts do not pose a threat to institutional analysts at this time, however, they provide a more accessible information basis and improve decision-making for individual investors.

Keywords: Social Investment Platforms, Social Media, Crowd Analysts, Financial Analysts, Natural Language Processing.

1   Introduction

In this paper, we examine the information dissemination role of financial analyst reports made available by non-professional "crowd" analysts on social investment platforms compared to institutional reports, issued by professional sell-side financial analysts. The number of crowd analyst reports has increased in recent years and research has only just started to investigate these information intermediaries [1-3]. We investigate how content and expressed opinion of crowd and institutional analyst reports are related to each other. We also investigate to what extent and how fast both report types incorporate up-to-date information.

The emergence of crowd analysts is a relatively new phenomenon, creating "additional content that adds to or otherwise affects the information content of firm disclosures […] as a result of changes in technology and the media" [4]. Their analyses are made available to other market participants via social investment platforms. These

platforms allow crowd analysts to publish their reports, analyses, interpretations, or recommendations. In contrast to institutional analyst reports, the information published on these platforms is also available to non-professional investors who cannot afford institutional reports, as subscriptions often cost several thousand dollars per user [5-7].

We therefore ask the following overall research question: What is the role of crowd analysts within the market of financial information intermediaries? Previous studies have come to different conclusions on this question. While Drake et al. [8] see crowd analysts as a threat to the business of institutional analysts, Kommel et al. [9] cannot confirm this. In contrast to these prior studies, our study examines both kinds of reports (institutional and crowd) on a textual level. This will allow us to gain a deeper insight into the kind of content these two report types bring to the market. This sheds light on the informational contribution crowd reports can provide for investors. With this study, we also contribute to the literature of the changing environment of financial analysts in general, as crowd analysts emerge as a new phenomenon in the age of social media and platform services.

We analyze 7,836 company-related analyst reports from a social investment platform of all companies listed in the Dow Jones Industrial Average (DJIA) between 2015 and 2019. These reports are compared to 24,606 institutional reports for the same companies and time period. Further, we use 730 conference call (CC) transcripts of these companies to identify important news keywords discussed in the CCs and examine whether institutional and crowd analysts took up these keywords. For examining similarities between the institutional and crowd analyst reports, we use TF-IDF-based cosine similarity [10]. Our empirical results show that crowd analysts provide similar information as institutional analysts, however, with a time lag of a few days.

This paper is structured as follows: Firstly, we provide a theoretical background on traditional financial analysts and crowd analysts. Based on this, we develop research questions and hypotheses and explain the methodological background. This is followed by a description of the dataset and its pre-processing. A detailed description of our analysis and our empirical results are presented afterward. Within the discussion section, we provide further interpretations of our results. The paper closes with a conclusion and an outline of possible directions for future research.

2 Theoretical Background

2.1 Institutional Analysts

According to New Institutional Economics, the existence of financial analysts is justified by the demand of information intermediaries reducing information asymmetries between market participants, precisely between managers possessing insider information and investors without access to this information [11], [12]. The role of traditional financial analysts has been examined in depth by existing literature. In their role as information intermediaries, they create information by discovery and interpretation. They reduce the asymmetry by dissminating information, thus making it

available to (potential) investors and reducing information asymmetry [13-15]. This role is particularly relevant at times when companies publish financial earnings and it reduces the time to information incorporation at the financial markets, which in turn improves market efficiency [13], [14], [16]. Sell-side institutional analysts are often employed in brokerage firms, research institutes, or investment banks. Brokerage firms and research institutes are usually commissioned to produce analyst reports [17].

Institutional analyst reports are characterized by analyzing information on the financial and earnings position of companies and macro- and microeconomic factors and pass on information interpretations in order to facilitate better decisions [18]. Assumingly, financial analysts have privileged access to non-public company information, why their information is considered particularly relevant [14]. A traditional analyst report contains an earnings forecast, a stock price target, and a recommendation about buying, holding or selling the financial instrument, as well as arguments to support the recommendation [18]. A substantial share of all analyst reports is published in direct conjunction or shortly after a firm's CC, often adopting and disseminating the CCs' content and providing a related interpretation [19]. CCs are quarterly meetings of the firm's management and analysts to discuss the firm's development and answer questions of analysts.

2.2 Crowd Analysts

Similar to institutional analyst reports, crowd analyst reports usually provide investors with an earnings forecast and a recommendation about the company's stock, fulfilling an information dissemination role [3]. The main difference to traditional analyst reports lies in the audience, that are, mainly private investors. Most of the analysts providing reports on investment platforms are non-professionals. The author collective of social investment platforms also contains investment professionals and experienced individuals from the financial sector conducting the research in their free time [20].

In recent years, the literature has started to investigate the phenomenon of crowd analysts. Chen et al. [1] find that non-institutional crowd analyst reports can be used to predict stock price developments and earnings surprises. Similar results can be found for the crowd's earnings forecasts, even stronger when the contributing crowd is larger [2]. The existence of crowd reports also helps investors to mitigate a negative bias in institutional reports, improving the prediction of earnings surprises [21]. A recent study supports the growing relevance of crowd analyst, finding that bearish recommendations provide more accurate stock price prediction than recommendations in traditional analyst reports [3]. Campbell et al. [22] find that stock markets react with a price increase to articles with a positive tone, indicating their credibility. Farrell et al. [6] focus on the benefit for individual investors, that are provided with more and accessible information through the social investment platform, decreasing the information advantage compared to professionals, while liquidity on financial markets increases. This aspect can be supported by easier readability of crowd reports that at the same time provide a higher information density, potentially leading to lower costs of information processing [23]. Another strand of literature has examined the relationship between crowd analysts and institutional analysts. Crowd analysts and their confirmed effect on

the accessibility of information to non-institutional investors have the potential of disciplining traditional analysts by lowering the incentive to publish pessimistic and too conservative short-term earnings forecasts [24]. The authors find the forecasts being more optimistic yet accurate. They also find crowd earnings forecast to be published much later than earnings forecast from their institutional peers. Drake et al. [8] investigate crowd analyst reports and their findings suggest a competitive threat through pre-empting traditional analysts' reports.

2.3   Research Question and Hypothesis Development

In the previous literature on crowd analysts, the main focus is dedicated to the evaluation of crowd analyst reports' accuracy and performance [1], [3]. Comparative studies that consider institutional and crowd analyst reports are rare and provide mixed evidence [8], [9]. These studies essentially compare the sentiment of crowd analyst reports with price forecasts from analyst databases (e.g., I/B/E/S). However, these studies cannot determine what information is provided by these groups of analysts and to what extent interdependencies exist between these groups in content and expressed opinion. Comparing crowd analysts and institutional analyst on a textual level has not been extensively covered in research. To close this gap, we compare the text contents provided by them. This is crucial for a better understanding of the role of crowd analysts in relation to their professional peers. Because after all, it is the text that analysts use to communicate their findings to the capital market.

To answer our overall research question, we split it into two sub-questions. As analysts function as information intermediaries and information discovery is one of their primary roles [19], the timely supply of relevant information to investors has to be fulfilled. This leads to RQ1.

  RQ1: How does the capability of timely information discovery, creation and dissemination distinguish between institutional and crowd analysts?

Besides reporting in a timely manner, reporting new information is another element of the information discovery role, leading to RQ2:

  RQ2: To what extent do institutional and crowd analysts provide related content and similar opinions?

Aspects such as a possible closer relationship to firm management and greater resources regarding financials and information processing possibilities establish a privileged access on the side of institutional analysts [19], [25], [26]. We assume that institutional analysts can analyze and publish new information faster than crowd analysts and therefore, contribute more to the reduction of information asymmetry. These advantages would justify the existence of institutional analysts in the context of New Institutional Economics [12], even though low-priced or free alternatives are made available by crowd analysts. We assume the topicality of institutional analyst reports to be higher and formulate the following hypothesis addressing RQ1.

  H1: Institutional analysts provide more topical information to investors compared to crowd analysts.

Regarding RQ2 we assume a high degree of similarity in content and opinion between reports of crowd and institutional analysts. However, we assume that crowd reports are

more related to preceding institutional reports than institutional reports to preceding crowd reports. As crowd analysts have fewer resources for information retrieval and information processing compared to their professional peers mostly employed by international brokerage companies, they have a strong incentive to rely stronger on the research conducted by institutional analysts and, therefore, disregard their own content and opinions.

H2.1: The originality of crowd reports content is lower than that of institutional reports.

H2.2: The originality of crowd reports opinion is lower than that of institutional reports.

3 Research Methodology

To compare the similarity of the reports' content, we use cosine similarity as a widely used approach in accounting and finance contexts to analyze documents of financial communication, e.g., analyst reports [27], financial product descriptions [28] or annual reports [29]. We apply this measure on a TF-IDF (term frequency – inverse document frequency) document representation [30]. The cosine similarity (1) is calculated between the word vectors A and B for each document pair. The cosine similarity is a measure for the angle between the vectors A and B [31]. The score can take a value between 0 and 1, while a high similarity score indicates a higher similarity between the two documents. The cosine similarity is especially useful for the comparison of sparse vectors (vectors containing many zero values) because it is robust against the extension of vectors by more zero values [31]. Since vectors of a term-document-matrix are typically very sparse, the cosine similarity is suitable for our application. The combination of cosine similarity and TF-IDF has proven to be a good measure for detecting documents containing new information in the area of novelty detection [32]. Since we are confronted with a very similar problem, we apply this measure.

$$\text{Cosine Similarity } (A, B) = \frac{\text{Cross product } (A, B)}{\sqrt{\text{Cross Product}(A) * \text{Cross Product}(B)}} \tag{1}$$

To evaluate the opinion addressed within the reports, we use sentiment. In the context of finance and accounting research, measuring the sentiment provides insights on how the author of a document perceives corporate information such as financial news [33], annual reports [34], or analyst reports [35]. We use a dictionary-based approach, assigning each word within a document a positive, negative, or neutral connotation [36]. We apply the Loughran/McDonald positive and negative word lists developed for finance-related documents [37]. The sentiment score of a document can take a value between –1 and 1.

$$Sentiment\ Score = \frac{n\ positive - n\ negative}{n\ positive + n\ negative} \tag{2}$$

4 Dataset and Descriptive Statistics

Our dataset is compiled from three sources. Institutional analyst reports as well as CC transcripts are obtained from Refinitiv Thomson ONE and the crowd analyst reports are from an online platform providing crowd equity research. The observation period of four years ranges from 07-01-2015 to 06-30-2019. To ensure that the observed companies are sufficiently covered by both professional and crowd analysts, we have selected the 31 companies that have been a constituent of the DJIA during our observation period as a sample. Our sample consists of 24,606 institutional and 7,836 crowd analyst reports written in the English language. Only reports are selected that cover one individual company exclusively. We consider the transcripts of 482 CCs that took place during the observation period. In addition, 248 CC transcripts taking place before the start of our investigation period have been indirectly included in our analysis as reference transcripts (further outlines on this in the analysis section).

To prepare the documents for further analysis, we apply standard pre-processing methods. We follow Huang et al. [19] and removed any boilerplate, disclaimer, tables and graphs from the analyst reports. From conference call transcripts we separately extract metadata (e.g., who is speaking) and content-related data (transcript of the spoken word). As the conference call transcripts have a clearly defined structure, i.e., the metadata and the actual content is consistently separated by the same text pattern, the separation between meta- and content-related data is done by applying a simple rule-based string processing. For all document types, we drop punctuation, figures, and non-ASCII characters and transform the text to lower case. Utilizing *gensim* phrase detection allows us to concatenate common multi-word expressions (e.g. cash flow -> cashflow). The text is tokenized to unigrams and then stemmed utilizing the Porter stemmer [38]. We further drop the respective company name and security ticker as well as stopwords (e.g. "and", "the"), and words with one or two letters, as these words will most likely not add actual content.

To get a better understanding of analysts' information, we analyze their research output. In Figure 1, research output is plotted against the time relative to the companies' CC. The left plot is showing the number of reports, whereas the right plot is showing their length. Huang et al. [19] highlight the importance of CCs for institutional analysts. They found that most reports are published on the (following) day of the companies' CC. Our data confirm this observation. For crowd analysts, we see a similar pattern. However, the timeframe of increased publication activity is considerably broader. As the research output of institutional analysts drops to the normal level only four days after the CC, we observe increased publication activity until ten days after the CC for crowd analysts. Furthermore, the crowd analysts are less focused on the CC, as they publish relatively more reports between CCs than professional analysts.

The evidence from report length (after described removal of boilerplate and disclaimer) shows the inverted case. Reports published close to the CC are considerably shorter than reports published between conference calls. This effect is stronger for institutional reports, which are longer in general. This finding is less surprising, as analysts, that want to publish their reports on the day of the CC, are faced with notable time constraints.
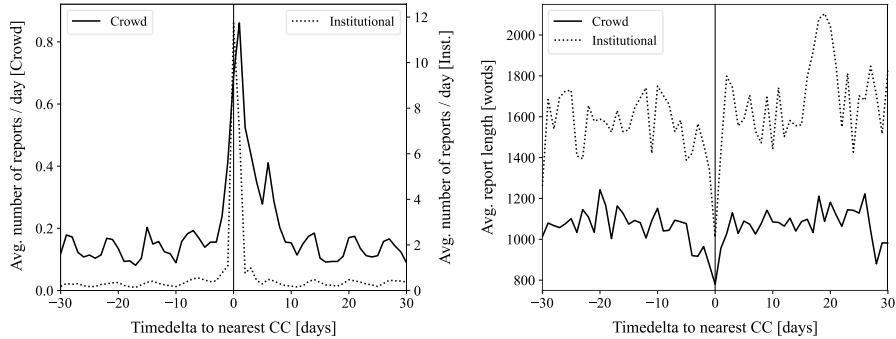
Figure 1. Publication pattern and report length around CCs

5 Analysis

5.1 Information Topicality

To answer RQ1, we first have to identify analyst specific news. One approach would be to use public news streams and filter for news related to the corresponding company. This would give us a comprehensive collection of news, but it would not tell us whether a certain news item is important for shareholders, and thus for analysts. We would also oversee news, which might have an impact on the company but where the company name is not mentioned in the news article (e.g., macroeconomic or political events).

To overcome this issue, we chose an indirect approach to extract relevant news. We compare the corpus of CCs and extract words that have been discussed substantially within a CC (mentioned five times or more) but were not mentioned within the last eight CCs (two years). The CCs are usually held in the context of quarterly financial reports [39]. These words could either describe news that emerged between the current and the last CC or new information that is just released by the management. The last eight CCs were chosen to get a reference corpus that is comprehensive enough to filter words usually discussed within a CC. For this reason, we also used the transcripts of CCs conducted prior to the observation period's beginning. As the timeframe of the reference corpus ranges over two years, seasonal influences are prevented. The threshold of five mentions for the extracted keyword was determined after a manual review of the keyword lists. A low threshold results, especially in the extraction of misspelled words, whereas a higher threshold leads to important news being overlooked. A threshold of five balanced out these effects quite well. We also ran the analysis with different thresholds and the results remained robust.

To get a better understanding of the nature of the extracted keywords, we provide an example. From Apple's CC on the 1st of August 2017 the keyword "ARKit" was extracted, which referred to a platform for developing augmented reality applications previously announced by Apple during their 2017 developer conference on the 5th of June 2017. During the CC the ARKit was mentioned within the presentation and discussion section.

We only consider CCs happening from 12-31-2015 onwards (N=421) to ensure that enough reports being observed before the respective CC. However, our results remain stable when considering all CCs. For 264 (62,71%) of the remaining CCs we could identify at least one keyword. We assigned analyst reports to these CCs that cover the same company and have been published within a timeframe of 360 days around the CC. A single report might be assigned to multiple CCs. For each assignment, we checked whether the report contains at least one extracted keyword. If so, we labeled the specific CC/report combination as news adapted.

In Figure 2, the proportion of report/CC combinations with existing news adoption is plotted against the time difference of CC and report. For clarity, the plotted data is aggregated on a weekly interval. The solid vertical line indicates the CC the keyword was extracted from. Just after the previous CC (dashed-dotted vertical line) the adoption of these words into the reports increases, as the news start to become public.
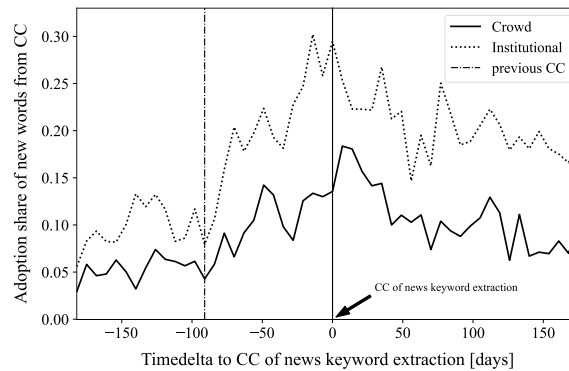

Figure 2. News adoption of crowd and institutional analysts

This is identical to the presented example, as Apple's ARKit was announced before the CC where it was discussed. The spread in news adoption between institutional and crowd analysts widens, as the institutional analysts are more likely to cover the news.

For the reports published during the day of the CC (solid vertical line) and the following six days, we see that news adoption for institutional reports peaked (29.47%). During this time period of highest analyst output, only 13.55% of the published crowd analyst reports covered the extracted news keywords. χ^2 test proves this difference to be highly significant (p<0.001). Crowd reports only reach their maximum news adoption in the second week after the CC (18.47%).

The results clearly show that institutional analysts can filter relevant news even before the CC from the continuous news stream to a greater extent than crowd analysts. This allows them to awaken investors' awareness regarding these topics, whereas crowd analysts take considerably longer and only reach their maximum news adaptation more than one week after the topic has already been discussed in the CC. However, the news adoption is by then still significantly lower than for institutional analysts before. This result corroborates H1 and further shows that institutional analysts satisfy their task of information discovery [19] better than crowd analysts.

5.2   Information and Opinion Originality

To evaluate the extent, crowd analysts provide similar information and opinion as institutional analysts and vice versa (RQ2), we compare each report with all reports of the opposing group published within a tight timeframe. To implement this approach, we build report pairs consisting of one institutional and one crowd analyst report as shown in Figure 3. Thereby, only pairs are formed that were published within the $PairingInterval$. The length of the $PairingInterval$ has to be specified. The cosine similarity (1) between the TF-IDF vectors of paired reports is calculated. For the similarity analysis, we reduce our feature set by excluding words occurring in more than 90% or less than 0.02% of all corpus documents. These cutoffs are useful to extract only words with high information content [40]. The upper cutoff of 90% is applied to exclude very frequent words that do not add information to the text but are not already filtered out as stopwords. The lower cutoff of 0.02% (equivalent to six reports) filters especially wrong spelled words.
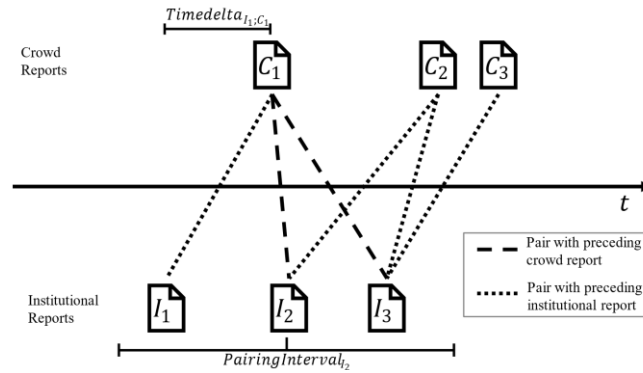

Figure 3. Building of report pairs

The sample of report pairs is divided into pairs with preceding institutional and preceding crowd reports. Within each subsample, we group the pairs by their time difference (number of days between publication of paired reports). For each group, the mean cosine similarity was calculated and plotted in Figure 4. The error bars indicate the 95% confidence interval of means. The highest similarity can be observed for same-day report pairs, which is not surprising. For this kind of report pairs, we cannot determine whether the institutional or the crowd report is proceeding. When looking at pairs with a time difference of one day or more, we observe an interesting pattern. For pairs with preceding crowd report (dashed line) we observe a steep decline in similarity just from the time difference of one day. However, if the institutional report was published first, the similarity remains relatively high up to a time difference of five days. For report pairs with larger time differences, the similarities of both groups are aligned again, and the effect is strongly reduced. This indicates that crowd reports tend to refer more to institutional reports than vice versa.
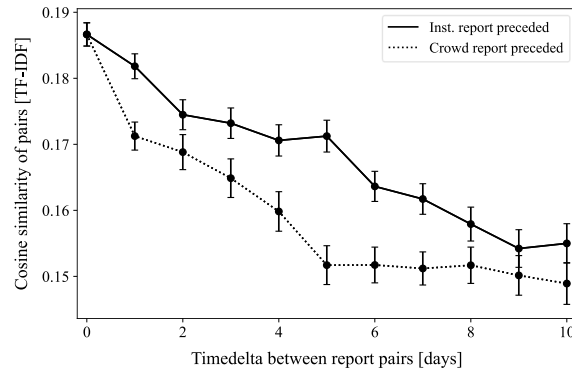
Figure 4. Originality of content

In order to deepen this analysis, we look at all report pairs together, which have a time difference between one and ten days (*PairingIntervall*: [-10;-1 & +1;+10]). Report pairs published on the same day cannot be considered, as it is not possible to determine which report was published first. As already discussed and derived from Figure 4, it is inappropriate to include report pairs with very long time differences as the effect is mainly observed between report pairs with a few days time difference. When combining all these report pairs as mentioned above, we get 71,011 report pairs with preceding institutional and 54,807 pairs with preceding crowd report. The mean cosine similarity across pairs with preceding institutional (crowd) report is 0.1684 (0.1585). The delta of 0.0099 is highly significant (p<0.001). This gives evidence for H2.1 being institutional analyst reports more original in content than their non-professional peers. To provide an intuition for the absolute level of cosine similarity of 0.1684 and 0.1585 respectively, we compare it to the average similarity of report pairs within each report type (crowd and institutional). For crowd (institutional) reports the cosine similarity amounts to 0.2301 (0.2206). We already discussed the importance of the CC. To refine our analysis, we divide our sample into pairs published close to the CC (CC-timeframe: ranges ten days prior to ten days past the CC) and pairs published outside of this interval (Non-CC-timeframe). Pairs extending over both timeframes are excluded. Within the CC-timeframe, we observe an overall higher similarity (Table 1). This is not surprising, as the CC expose both groups to similar information.[1] The delta between the groups (inst. preceding / crowd preceding) is however 96% larger during Non-CC-timeframes. This indicates that crowd analysts make relatively greater use of information from their institutional peers in times of low information density.

---

[1] The strong influence of the conference call on the content of the reports is also evident when calculating the average similarity between the reports and the conference call transcript for this period. Pairings between conference calls and crowd reports show a similarity of 0.2706 whereas the pairings between conference calls and institutional reports show similarities of 0.2912. These values are higher than the similarity between crowd and institutional reports.

Table 1. Cosine similarity between different group report pairs

Sample	Group	Cosine similarity	N
	Inst. preceding	0.1684	71,011
Overall	Crowd preceding	0.1585	54,807
	Delta	*0.0099****	
	Inst. preceding	0.1711	44,615
CC-timeframe	Crowd preceding	0.1643	30,493
	Delta	*0.0068****	
	Inst. preceding	0.1651	22,555
Non-CC-timeframe	Crowd preceding	0.1518	19,813
	Delta	*0.0133****	
*** p <0.001			

In addition to the comparison of content, we also compare the authors' opinions expressed within the document. Based on the finance-related sentiment dictionary of Loughran and McDonald [37], we count the number of positive and negative words within each document and calculate the documents' sentiment polarity by applying (2). On average, we observe a more positive sentiment of institutional reports (mean polarity: +0.017) compared to crowd reports (mean polarity: -0.060). This finding is in line with the comprehensive literature on institutional analyst optimism [41].

We use the same matching applied for content comparison. To evaluate whether both analyst groups have a similar opinion regarding a specific company during a specific point in time, we calculate the Pearson correlation between the sentiment polarities of matched reports. We find a highly significant positive correlation between the sentiment polarity of crowd and institutional reports for all subsections (Table 2). Since we are interested in whether the institutional analysts or crowd analysts are opinion leaders, we examine the correlation coefficients' delta. Overall, pairs with preceding institutional reports have a significantly higher correlation, indicating institutional analysts to be opinion leaders. We apply Fisher-z transformation to evaluate the significance of the difference in correlation coefficients.

Conversely to the adaptation of content, we recognize that crowd analysts might be especially influenced by the opinion of institutional analysts during the CC timeframe. During this timeframe, the correlation coefficient is 57.1% higher for pairs with a preceding institutional report (0.1904) compared to pairs with a preceding crowd report (0.1212). Outside of the CC timeframe, we observe only a small, insignificant delta in favor of the professional analysts. To make sure that the result from CC-timeframe is not purely driven by the sentiment conveyed within the CC, we controlled for the CC's

sentiment polarity by applying partial correlation [42].[2] Our results remain robust and the delta of the correlation coefficients during the CC time span remains significantly positive. Moreover, the adaptation of the opinion from the CC is significantly higher for institutional analysts (r=0.2374) than for crowd analysts (r=0.0635). These values are not based on the report-to-report pairing used above but by mapping the reports from within the CC-timeframe against the corresponding CC. If only reports published after the CC are considered, the correlation for crowd reports increased slightly but the correlation of institutional reports is remaining the same.

Table 2. Correlation of sentiment polarity between different group report pairs

Sample	Group	Correlation (Pearson)	Partial correlation†	N
Overall	Inst. preceding	0.1646***		71,011
	Crowd preceding	0.1087***		54,807
	Delta	*0.0559****		
CC-timeframe	Inst. preceding	0.1904***	0.1680***	44,615
	Crowd preceding	0.1212***	0.1220***	30,493
	Delta	*0.0692****	*0.0460****	
Non-CC-timeframe	Inst. preceding	0.1065***		22,555
	Crowd preceding	0.0924***		19,813
	Delta	*0.0141*		

*** p <0.001 † Partial correlation controls for the sentiment of the CC.

Our results suggest that the opinion of institutional analysts might influence crowd analysts during times of high information density. Thus, we can confirm H2.2 for the timeframe of the CC. During times of low information density, they rather form their own opinion. This is in line with the Social Impact Theory proposed by Latané [43], which states that the crowd size is positively related to crowd influence. The evidence of low adoption of CC sentiment by crowd analysts compared to institutional analysts can be attributed to the fact that the extraction of information from an analyst report appears to be much easier than the information processing of a CC transcript.

6   Discussion

Our results clearly show that institutional analysts are still intermediaries ensuring the timely publishing of new information. These findings are in line with Jame et al. [24], finding a delay of crowdsourced earnings forecasts. Crowd reports lack significantly in the timely provision of relevant news. This indicates the high relevance of institutional

---

[2] The correlation of the sentiment scores of the paired institutional and crowd reports is calculated after the influence of the sentiment from the conference call is eliminated from both variables. The partial correlation can be implemented by regressing the sentiment scores first from the crowd and second from the institutional reports against the conference call sentiment and then calculating the correlation between the residuals of these two regressions.

analysts for information dissemination, reduction of information asymmetry, and ensuring efficiency on capital markets. Not only do the results show a timelier adoption of news from the CC, but also the capability to identify relevant information before a CC. For crowd analysts the adoption of news is significantly lower, relevant information is reported later after it was already confirmed within the CC. Therefore, institutional analysts fulfill the function of information discovery [19] better than crowd analysts. To answer our first research question, it is apparent that crowd analysts take more time than institutional analysts to fulfill the information dissemination function. A reason can be the lack of resources, such as time and accessibility, or delayed quality control mechanisms of the platforms in opposition to institutional providers that aim to publish their services as soon as possible, while institutional analysts receive privileged access to information. A potential disrupting influence towards the financial analyst business cannot be confirmed in the question of timeliness.

Addressing the second research question on originality of content, cosine similarity results suggest that crowd reports provide similar information as preceding institutional reports significantly more than institutional reports to preceding crowd reports within a short timeframe of ten days. The ratio converges for longer timeframes. Referring to the analysis on timeliness of the reports, the results hint to crowd analyst reports not only being delayed in adopting and disseminating information but also relying on institutional analyst reports as an information basis. The division into two timeframes shows that this effect is lower during times of the CC and higher between CCs. In times of low information density when the firms provide no information, crowd analysts rely more on institutional analysts than in high information density times. Lower information availability outside of CC timeframes leads to higher costs of information procurement, incentivizing crowd analysts with fewer resources to rely on content previously created by institutional analysts.

Examining the opinion through sentiment polarities, we find that both crowd and institutional analysts adopt the sentiment of the CC to a large extent. This effect is stronger for institutional analysts. A reason may lie in the possibility of institutional analysts attending the CCs and contribute to shaping the opinion [44]. Not only in content but also in opinion, we observe crowd analysts adopting the interpretational tone of institutional analysts. This result is strong for the high information density timeframe. Interestingly, this observation cannot be made in times of low information density, indicating the creation of original opinions by crowd analysts. Another reason can also be attributed to the fact that information extraction from analyst reports appears to be easier than from CC transcripts. Our findings clearly show that institutional investors are still leading in content and opinion compared to crowd analysts, even though lower-priced or free alternatives are available to investors. This justifies the existence of institutional analysts in the context of New Institutional Economics [12].

Our study is subject to some limitations. To ensure appropriate coverage among crowd analysts and institutional analysts on the sample companies, we are restricted to an equity index with a rather small number of companies. We use a TD-IDF document representation to apply cosine similarity analysis on analyst reports. Other document representation, especially topic models, might enhance interpretability and add further assumptions and complexity to the analysis. Alternatively, mean word embeddings

(e.g., word2vec or GloVe) or document embeddings (e.g., doc2vec) could be used as text representation. As a robustness check, we performed our analyses using meaned word embeddings based on pre-trained GloVe embeddings [45]. Thereby each document is represented by a 300-dimensional vector. The basic structure of the results remains stable.[3] The advantage of word embeddings is that the semantic similarity of different words is considered. In the area of novelty detection, however, the loss of word specificity in word embeddings based measures leads to underperformance compared to TF-IDF-based similarity measures on novelty detection tasks [46]. Numerous alternatives to the cosine similarity are available, but this measure's effectiveness has been demonstrated in practical applications despite its limited theoretical foundation. Moreover, it is less sensitive to document length than, for example, the Manhattan distance [32]. For this reason, we consider it appropriate to calculate the similarity of documents based on TF-IDF in combination with cosine similarity.

For the sentiment analysis, we decided to use a dictionary approach designed for financial contexts, widely used in analyst report research. It has no need for labeling that could be affected by the subjective opinions of the person conducting the labeling. However, for text mining in analyst reports, other approaches such as a naïve Bayes approach have been assessed as more accurate [35]. Furthermore, we cannot rule out the possibility that professional analysts are also enrolled on the equity research online platform and we, therefore, allocate institutional analysts' ability to a certain extent to the abilities of crowd analysts.

7   Conclusion

In this paper, we examine the information dissemination role of financial analyst reports made available by non-professional "crowd" analysts on social investment platforms compared to institutional reports issued by traditional financial analysts. In recent years, the number of institutional analysts is decreasing [47], whereas platform business models and social media are constantly growing. Non-institutional analyst reports are available for a wider range of market participants, especially individual investors, and therefore allowing a better basis for decision-making in financial markets. We examine institutional and crowd reports from 2015 to 2019 concerning their capability of dissemination new information derived from CCs and their similarity to each other. We find that institutional analysts are faster in disseminating news and relevant information. Leading in topicality, institutional analysts mainly use CCs for their analysis, while crowd analysts tend to rely on institutional analyst reports as an information source. This effect is more pronounced in times of a low information density between CCs. We also find that crowd analysts are influenced by the opinion expressed in institutional analyst reports during the time of the CC. In times of low information density, they disseminate a more individual opinion.

---

[3]   The results of the sensitivity check are not included into this document but are available upon request.

Our study provides evidence on the role of crowd analysts. First of all, findings on a more topical news adoption from institutional analysts (RQ1) indicate possible incentives for crowd analysts to follow the content and opinion of institutional analysts rather than conducting their own research. This presumption is precisely confirmed when looking at the relation of content and opinion (RQ2) between crowd and institutional analysts. Since the observed delay is only a few days, this does not mean that crowd reports are worthless. Rather, it shows that the vehicle of crowd reports can provide information to investors that is otherwise only available to institutional investors with high research budgets. The delay might be less serious for investors with long-term investment horizons. For investors with high investment volumes and short-term investment horizons, it seems reasonable to continue relying on expensive institutional reports despite the low-cost alternative of crowd reports.

Our research contributes to the literature on the role of crowd analysts and the value they can provide to market participants through social investment platforms in contrast to institutional analysts. We also provide a deeper understanding of crowd analysts' role within the capital market for individual investors, institutional analysts, researchers and regulators. Social investment platforms can use these results to derive measures on how to improve their information creation processes and objectives on how to become more independent from institutional business research. Crowd analysts should be encouraged to search for private information to create additional value for market participants. We also provide an approach on how to extract relevant keywords from documents such as CCs without requiring a topic modeling approach. The results indicate that even though crowd analysts currently do not pose a threat to the market position of traditional analysts, there is some potential to grow in relevance, especially for less sophisticated and non-institutional investors. Through better accessibility and easier information processing of crowd reports for market participants, crowd analysts might shape the market of business research in the future.

References

1. Chen, H., De, P., Hu, Y.J., Hwang, B.-H.: Wisdom of crowds: The value of stock opinions transmitted through social media. The Review of Financial Studies 27(5), 1367–1403 (2014).
2. Jame, R., Johnston, R., Markov, S., Wolfe, M.C.: The Value of Crowdsourced Earnings Forecasts. Journal of Accounting Research 54(4), 1077–1110 (2016).
3. Jin, Y., Ye, Q., Gao, C., Xia, H.: The Value of Amateur Analysts' Recommendations Extracted from Online Investment Communities. In: PACIS 2019 Proceedings (2019).
4. Miller, G.S., Skinner, D.J.: The Evolving Disclosure Landscape: How Changes in Technology, the Media, and Capital Markets are Affecting Disclosure. Journal of Accounting Research 53(2), 221–239 (2019).
5. Brush, S., Spezzati, S.: How do you put a price on investment research?, https://www.bloomberg.com/professional/blog/put-price-investment-research-2/ (Accessed: 11-27-2020)
6. Farrell, M., Green, T.C., J., R., Markov, S.: The Democratization of Investment Research: Implications for Retail Investor Profitability and Firm Liquidity. Available at SSRN 3222841, (2018).

7. Gomez, E., Heflin, F., Moon, J., Warren, J.: Crowdsourced financial analysis and information asymmetry at earnings announcements. Georgia Tech Scheller College of Business Research Paper, (2018).
8. Drake, M.S., Moon, J., Twedt, B.J., Warren, J.: Are Social Media Analysts Disrupting the Information Content of Sell-Side Analysts' Reports? Available at SSRN 3456801, (2019).
9. Kommel, K.A., Sillasoo, M., Lublóy, Á.: Could crowdsourced financial analysis replace the equity research by investment banks? Finance Research Letters 29, 280–284 (2019).
10. Tata, S., Patel, J.M.: Estimating the selectivity of tf-idf based cosine similarity predicates. SIGMOD Rec. 36(2), 7–12 (2007).
11. Frankel, R., Li, X.: Characteristics of a Firm's Information Environment and the Information Asymmetry between Insiders and Outsiders. Journal of Accounting and Economics 37, 229–259 (2004).
12. Rischkowsky, F., Döring, T.: Consumer Policy in a Market Economy: Considerations from the Perspective of the Economics of Information, the New Institutional Economics as well as Behavioral Economics. Journal of Consumer Policy 31, 285–313 (2008).
13. Chen, X., Cheng, Q., Lo, K.: 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 (2010).
14. Frankel, R., Kothari, S.P., Weber, J.P.: Determinants of the Informativeness of Analyst Research. Journal of Accounting and Economics 41(1–2), 29–54 (2006).
15. Lawrence, A.: Individual investors and financial disclosure. Journal of Accounting and Economics 56(1), 130–147 (2013).
16. Elgers, P.T., Lo, M.H., Pfeiffer, R.J.J.: Delayed Security Price Adjustments to Financial Analysts' Forecasts of Annual Earnings. The Accounting Review 76(4), 613–632 (2001).
17. Brown, L.D., Call, A.C., Clement, M.B., Sharp, N.Y.: Inside the "Black Box" of Sell-Side Financial Analysts. Journal of Accounting Research 53(1), 1–47 (2015).
18. Asquith, P., Mikhail, M.B., Au, A.S.: Information Content of Equity Analyst Reports. Journal of Financial Economics 75(2), 245–282 (2005).
19. Huang, A.H., Lehavy, R., Zang, A.Y., Zheng, R.: Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach. Management Science 64(6), 2833–2855 (2018).
20. Chen, H., Hu, Y.J., Huang, S.: Monetary Incentive and Stock Opinions on Social Media. Journal of Management Information Systems 36(2), 391–417 (2019).
21. Schafhäutle, S., Veenman, D.: Crowdsourced Earnings Expectations and the Salience of Sell-Side Forecast Bias. Available at SSRN 3444144, (2019).
22. Campbell, J.L., DeAngelis, M., Moon, J.R.J.: Skin in the Game: Personal Stock Holdings and Investor's Response to Stock Analysis on Social Media. Review of Accounting Studies 24(3), 732–779 (2019).
23. Palmer, M., Bankamp, S., Muntermann, J.: Institutional versus Crowdsourced Analyst Reports: Who Puts it in a Nutshell? In: PACIS 2020 Proceedings (2020).
24. Jame, R., Markov, S., Wolfe, M.: Does Crowdsourced Research Discipline Sell-Side Analysts? Available at SSRN 2915817, (2019).
25. Clement, M.B.: Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? Journal of Accounting and Economics 27(3), 285–303 (1999).
26. Clifton Green, T., Jame, R., Markov, S., Subasi, M.: Access to management and the informativeness of analyst research. Journal of Financial Economics 114, 239–255 (2014).
27. Eickhoff, M., Muntermann, J.: They Talk but What do they Listen to? Analyzing Financial Analysts' Information Processing using Latent Dirichlet Allocation. In: PACIS 2016 Proceedings (2016).

28. Hoberg, G., Phillips, G.: Text-Based Network Industries and Endogenous Product Differentiation. Journal of Political Economy 124(5), 1423–1465 (2016).
29. Lang, M., Stice-Lawrence, L.: Textual analysis and international financial reporting: Large sample evidence. Journal of Accounting and Economics 60(2-3), 110–135 (2015).
30. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Information Processing & Management 24(5), 513–523 (1988).
31. Han, J., Kamber, M., Pei, J.: Getting to Know Your Data. In: Han, J., Kamber, M., Pei, J. (eds.) Data Mining (Third Edition), pp. 39–82. Morgan Kaufmann, Boston (2012).
32. Zhang, Y., Callan, J., Minka, T.: Novelty and Redundancy Detection in Adaptive Filtering. In: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 81–88. Association for Computing Machinery (2002).
33. Liebmann, M., Hagenau, M., Neumann, D.: Information Processing in Electronic Markets measuring Subjective Interpretation Using Sentiment Analysis. In: ICIS 2012 Proceedings (2012).
34. Huang, X., Teoh, S.H., Zhang, Y.: Tone Management. The Accounting Review 89(3), 1083–1113 (2014).
35. Huang, A.H., Zang, A.Y., Zheng, R.: Evidence on the Information Content of Text in Analyst Reports. The Accounting Review 89(6), 2151–2180 (2014).
36. Loughran, T., McDonald, B.: The Use of Word Lists in Textual Analysis. Journal of Behavioral Finance 16(1), 1–11 (2015).
37. Loughran, T., McDonald, B.: When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. The Journal of Finance 66(1), 35–65 (2011).
38. Porter, M.F.: An algorithm for suffix stripping. Program 14(3), 130–137 (1980).
39. McKay Price, S., Doran, J.S., Peterson, D.R., Bliss, B.A.: Earnings conference calls and stock returns: The incremental informativeness of textual tone. Journal of Banking & Finance 36(4), 992–1011 (2012).
40. Patwardhan, S., Pedersen, T.: Using WordNet-based context vectors to estimate the semantic relatedness of concepts. In: Proceedings of the Workshop on Making Sense of Sense: Bringing Psycholinguistics and Computational Linguistics Together (2006).
41. Brown, L.D.: Earnings forecasting research: its implications for capital markets research. International Journal of Forecasting 9(3), 295–320 (1993).
42. Johnson, R.A., Wichern, D.W.: Applied multivariate statistical analysis. Prentice hall Upper Saddle River, NJ (2002).
43. Latané, B.: The psychology of social impact. American Psychologist 36(4), 343–356 (1981).
44. Cornaggia, J., Cornaggia, K.J., Xia, H.: Revolving Doors on Wall Street. Journal of Financial Economics 120(2), 400–419 (2016).
45. Pennington, J., Socher, R., Manning, C.D.: GloVe: Global Vectors for Word Representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014).
46. Wang, F., Ross, R.J., Kelleher, J.D.: Exploring Online Novelty Detection Using First Story Detection Models. In: International Conference on Intelligent Data Engineering and Automated Learning, pp. 107–116. Springer (2018).
47. Lee, J.: Analyst Jobs Vanish as a Perfect Storm Crashes Into Research, https://www.bloomberg.com/news/articles/2019-12-19/analyst-jobs-vanish-as-a-perfect-storm-hits-wall-street-research (Accessed: 11-27-2020)