The Relationship Between data privacy discourses and company performance at Facebook(2004–2021)

Dr.Prajwal Eachempati†
Trinity Business School
Trinity College
Dublin Ireland
prajwal.eachempati@adaptcentre.ie

Prof. Laurent Muzellec
Trinity Business School
Trinity College
Dublin Ireland
laurent.muzellec@adaptcentre.ie

Dr. Ashish Kumar Jha
Trinity Business School
University Trinity College
Dublin Ireland
akjha@tcd.ie

ABSTRACT

We use Facebook as a case study to investigate the complex relationship between the firm's public discourse (and actions) surrounding data privacy and the performance of a business model based on monetizing user's data. We do so by looking at the evolution of public discourse over time (2004–2021) and relate topics to revenue and stock market evolution. Drawing from archival sources like Zuckerberg LDA topic modelling is implemented to reveal 19 topics regrouped in 6 major themes. The paper aims to understand and put a value on the extent to which privacy disclosures tuned with specific topic keywords have a potential impact on the financial performance of social media firms. There we found significant relationship between the topics pertaining to privacy and social media/technology, sentiment score and stock market prices. Revenue is found to be impacted by topics pertaining to politics and new product and service innovations while number of active users is not impacted by the topics unless moderated by external control variables like Return on Assets and Brand Equity.

KEYWORDS

• public discourses; social media; topic modelling; privacy; business model; financial performance

1. Introduction

Data privacy is a controversial topic because it is an ambiguous construct. Data is managed and usually owned by digital platforms while privacy relates to the intimacy of users. Social media platforms in particular thrive on users' data. No other firm better than Facebook illustrates this tension between monetization and respect of users' data as well as the recurrent privacy crisis inherent to such business model (e.g. 2008, 2012, 2015, 2018, 2020). With 3 billion users, it is one of the world's largest aggregators of user data. According to Forbes, the Facebook brand is valued at $70.39 billion (fifth most valuable in the world) in 2020, generating $117,929 billion in revenue and $ 39,370 billion in net income in 2021 .

Yet, research related to data privacy has focused predominantly on user behavior and the related privacy paradox (Gerber et al., 2018). Few studies have looked at online privacy from the business perspective, with some notable exceptions. Stutzman et al.,(2011) and Pollach(2005) have explored the use of language by corporations in their official privacy policies and found that these formal policies can "obfuscate, enhance and mitigate unethical data handling practices" (Pollach, 2005, pp. 221). However, no studies have investigated the relationship between privacy discourse, public opinion crisis and the success of digital business as measured by Key Performance Indicators such as numbers of active users, revenue and stock market prices. This will help us to answer the following research questions:

-What is the corporate public discourse around privacy?

-What is the relationship between public discourse during privacy crisis and Key Performance Indications (active users, revenue and stock market prices?)

2. Data Collection and Research Methodology

The main motive of this research is to understand the thematic evolution of Facebook transcripts with time and how the themes related to privacy and business models are emerging. The detailed research methodology is illustrated below:

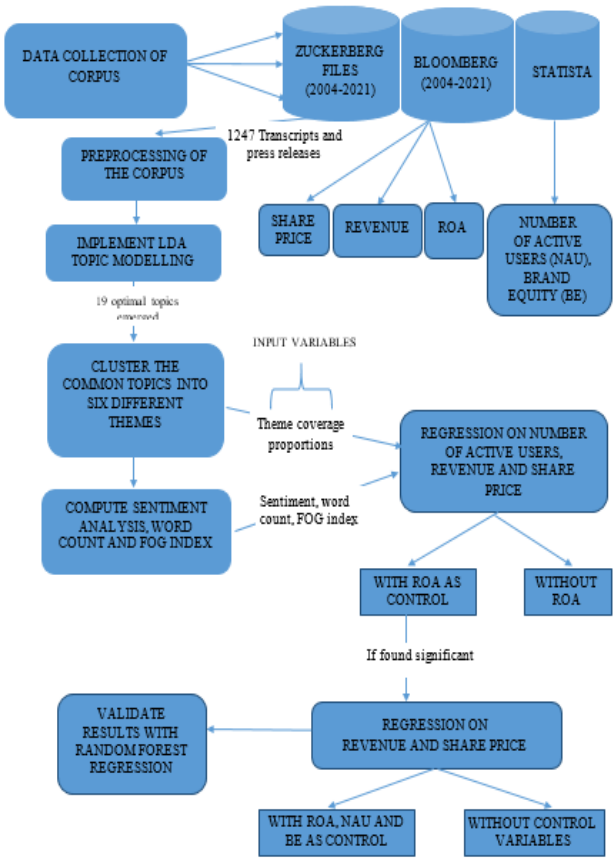

Figure 1. Research methodology

The Zuckerberg files between the years 2004 to 2021 are collected and filtered in terms of "Transcripts" to extract only the files, which are transcripts of Mark Zuckerberg's speeches. The search retrieved 1247 transcripts over the years. The LDA topic-modelling algorithm (Thielmann et al.,2021) was implemented to identify the major topics and themes covered in the transcripts. This would help understand which themes have evolved in importance for Facebook. Sentiment analysis was performed in R (Wang et al., 2022) to extract the sentiment of the Zuckerberg transcripts in order to understand how the announcements are strategically framed around issues of privacy and the impact on changing business models. The word count of each article is also computed and incorporated as a predictor in the dataset to understand how the degree of conciseness of a transcript/article would impact the performance of the firm. An additional metric to the degree of conciseness is also the rate of complexity and readability of the Zuckerberg announcements. This would provide an insight into how firms strategically use confounding and high complexity words without providing a clear assurance to the users about their stand on privacy. For this purpose, the metric FOG index (Ahmadi et al.,2021) is computed in R. After performing basic preprocessing, the transformed corpus is converted into a Document term matrix. A coherence score plot is constructed to find the optimal number of topics between 1 to 20 and the highest coherence score was found to occur at 19. The next use case of the algorithm was to map each document to its predominant topic cluster. Further, the top 15 frequently occurring terms in each topic were extracted to observe the context in each topic. Further analysis has revealed that there are six emergent themes among the 19 topics. The financial indicators of the firm performance measured by Share price, Revenue and Return on Assets are sourced from the Bloomberg database while the data on number of active users and brand equity is retrieved from the Statista database. The thematic topic coverage proportions, sentiment, word count and FOG index are regressed on the outcome variables Share price, Revenue and the number of active users with and without the impact of control variable "Return on Assets". Similarly, if the variables are found to be significant, brand equity is introduced as an additional control variable. This is to introduce the moderating influence of number of active users and brand equity on the outcome variables Share price and Revenue. The regression results are validated by the random forest regression machine-learning algorithm. The results are illustrated below:

Table 1. Summary of regression model results on Share price, Revenue and Number of Active users as outcome

Variables	Model 1	Model 2	Model 3
	Coefficients	Coefficients	Coefficients
governance	0.21(0.636)	-882(0.236)	-0.034(0.858)
community/ education	0.51(0.12)	-193.24(0.721)	-0.005(0.97)
politics	0.1 (0.83)	-1459.4(0.048)	0.522(0.011)
privacy	0.48*** (0.026)	-82.6(0.907)	-0.027(0.88)
product/ service/ innovation	0.08(0.882)	538.4 ***(0.003)	-0.002(0.992)
social media/ technology	0.074*** (0.42)	-545.3(0.358)	0.097(0.523)
Word count	0.17 (0.743)	3354 ***(0.0004)	0.185(0.412)

FOG index	0 (0.89)	-47.5***(0.004)	0(0.927)
Sentiment_ score	0.21*** (0.009)	-133.02(0.503)	0.07(0.191)

The three models Model 1, Model 2 and Model 3 imply the regression results for the key performance indicators Share price, Revenue and Number of active users respectively. For the first regression model with Share price as dependent variable, the topics pertaining to privacy, social media/technology and sentiment score are found to be the most significant positive drivers. This implies that providing assurance to users about privacy, latest technological trends and a more positive linguistic framing of transcripts has an impact on boosting the share price of Facebook. Similarly, for the second model with Revenue KPI as dependent variable, the news pertaining to politics, product/service/innovation news, word count and FOG index are found significant. This indicates that keywords related to political agendas and new products/services, the number of words in transcript and degree of complexity of transcript are important drivers of revenue. While, product and service-related news and word count are positive drivers, the news pertaining to politics and FOG index are negative drivers. This is because topics pertaining to politics may be controversial and detrimental to the reputation of Facebook in the wake of the Cambridge Analytica scandal and the use of complex words may cloud decision-making and cause a dip in revenue. However, political news is found to have a positive impact on the number of active users KPI as illustrated in Model 3. Overall, the topic-related variables privacy, social media/technology and sentiment score are found to be significant implying that the transcripts pertaining to privacy and social media and the overall sentiment have an impact on the key performance indicators of the firm.

Further, to improve the explainability of the model, other financial control variables like Profitability, Return on Assets and Price Earnings ratio have also been incorporated. However, Profitability and Price Earnings ratio were found to be highly correlated (>80%) and eliminated therefore, ROA is the only control variable considered.

Further, brand equity found to be significant is also incorporated as a control variable and the result is illustrated below for stock price and revenue as outcome variables respectively:

Stock price as outcome

```
Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                    -4.527e+01  4.773e+01  -0.948  0.34318
governance                      2.346e+00  4.975e+01   0.047  0.96240
community.education             1.510e+01  4.861e+01   0.311  0.75613
politics                       -2.931e+01  4.970e+01  -0.590  0.55545
privacy                        -6.009e+01  5.068e+01  -1.186  0.23614
product.service.innovation     -7.295e+01  5.243e+01  -1.391  0.16452
social.media.technology        -4.026e+01  4.708e+01  -0.855  0.39272
Word.count                      7.553e-04  2.633e-04   2.869  0.00423 **
FOG.index                       1.012e-02  4.752e-02   0.213  0.83137
Sentiment_score                 2.085e-02  4.586e+00   0.005  0.99637
Return.on.Assets                2.588e+00  3.440e-01   7.524 1.49e-13 ***
No.of.active.users.in.millions. 9.507e-02  3.523e-03  26.985  < 2e-16 ***
Brand.equity.in.millions...    -5.688e-05  1.279e-04  -0.445  0.65661
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 36.62 on 765 degrees of freedom
Multiple R-squared:  0.7553, Adjusted R-Squared: 0.75
F-statistic: 196.8 on 12 and 765 DF,  p-value: < 2.2e-16
```

Figure 2. Regression results for Stock price as outcome with ROA and Brand Equity as control variables

Word count, ROA and Number of active users are the only significant predictors of stock price with the effect of different topic-related news faded out as compared to the previous models with share price as KPI. Brand equity however, is not a significant control variable. This implies that brand reputation may not necessarily have an immediate impact on stock price unless moderated by a customer-centric variable.

This implies that brand reputation also boosts the revenue as it is moderated by an increase in number of active users.

REFERENCES

[1] Omid Ahmadi, Jacqueline Louw, Heta Leinonen, and Peter Yee Chiung Gan. 2021.Glioblastoma: assessment of the readability and reliability of online information. British Journal of Neurosurgery. 35, 5, 1-4. DOI: https://doi.org/10.1080/02688697.2021.1905772

[2] Anton Thielmann, Christoph Weisser, Astrid Krenz & Benjamin Säfken. 2021. Unsupervised document classification integrating web scraping, one-class SVM and LDA topic modelling. Journal of Applied Statistics DOI: https://doi.org/10.1080/02664763.2021.1919063

[3] Wei Wang, Lihuan Guo & Yenchun Jim. 2022. The merits of a sentiment analysis of antecedent comments for the prediction of online fundraising outcomes. Technological Forecasting and Social Change, 174. DOI: https://doi.org/10.1016/j.techfore.2021.12107.

[4] Pollach, I. (2005). A typology of communicative strategies in online privacy policies: Ethics, power and informed consent. Journal of Business Ethics, 62(3), 221-235. DOI: https://doi.org/10.1007/s10551-005-7898-3