

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

# Corporate Control? Measuring Private Sector Censorship of Social Media

---

A thesis presented in partial fulfillment of the requirements for  
the degree of

Master of Information Science in Information Technology

at Massey University,  
Albany, New Zealand



**MASSEY UNIVERSITY**  
**TE KUNENGA KI PŪREHUROA**  
**UNIVERSITY OF NEW ZEALAND**

Hanfeng Zhong

2016

(Supervisor: Professor Paul Watters)

## **Abstract**

Censoring activities on sensitive topics have played a significant role on social network sites (SNSs). Owing to the difference in politics, economics and cultures in the various countries, many social network sites including Facebook, Twitter, Google, Reddit and Imgur might implement different censorship standards according to the situation of the country. This study aims to explore whether governments' decision and censorship policies mentioned in previous studies have been implemented on main social network sites. Additionally, this article searches a list of sensitive keywords on each tested site, which is also the simplest approach applied to explore censorship on social network sites regulated using keywords filtering. Indeed, classifying a list of keywords into blacklist or merely blocking some defined sensitive topics refers to the primary method for censoring information on social network sites. The discussion makes us re-examine not only censorship on social network sites but also propose three possible conclusions concerning censorship on social network sites in specific country, such as 'censorship is weaker than we anticipated', 'some social network sites focus on supporting country's censorship' and 'censorship is imperfect to be implemented by social network sites'. As shown by results, some leaks still exist on current censorship of social network sites, while some sites fail to sensor harmful information that should be blocked. However, some harmless information is blocked by certain sites that may influence users' browse information. By analyzing the censorship data of blocked keywords and pornography sites on Facebook, Twitter, Google, Reddit and Imgur, this research highlights the defect of censorship implemented on social network sites.

**Keywords:** censorship standards, social network sites, censoring

## **Acknowledgements**

I appreciate all people who have provided me help and involved in my thesis. It would be impossible to finish my thesis without their help. Many people, including my supervisor, my school, my friends, and my family, have provided me useful suggestions in the data collection and thesis writing.

Firstly, I would like to say thanks to my supervisor, Professor Paul Watters, for his contribution to this thesis. He has helped me by providing useful materials, advice of good ideas, and instruction of data collection. The methodology was originally designed based on his suggestions. All in all, this thesis cannot be accomplished without his support. Then, I would like to thank my friends for helping me to eliminate my confusions. They gave me not only encouragement but also useful suggestions when I felt confused about how to continue with my thesis writing. Also, I would like to thank Massey University for providing us an excellent and academic study environment. By using Massey library, I can easily find out the academic papers I needed.

Finally, I appreciate for the supports of my parents. Even though they cannot speak English, they also gave me valuable suggestions for this thesis.

## Table of contents

Abstract.....	1
Acknowledgements .....	2
List of figures.....	4
List of Tables .....	5
Chapter 1: Introduction .....	6
Chapter 1.1: Research questions .....	10
Chapter 1.2: Hypotheses.....	11
Chapter 2: Literature review .....	13
2.1 Review of Facebook.....	13
2.2 Review of Twitter .....	19
2.3 Review of Google .....	23
2.4 Review of Reddit .....	30
2.5 Review of Imgur .....	33
Chapter 3: Methodology .....	36
3.1 Data collection.....	36
3.1.1 Data collection from Facebook.....	36
3.1.2 Data collection from Google.....	36
3.1.3 Data collection from Twitter .....	37
3.1.4 Data collection from Reddit.....	37
3.1.5 Data collection from Imgur .....	37
3.2 Measurement .....	37
3.3 Limitation of data collection.....	37
Chapter 4: Results.....	39
4.1 Measurement of Facebook censorship .....	39
4.2 Measurement of Google censorship .....	40
4.3 Measurement of Imgur censorship.....	42
4.4 Measurement of Reddit censorship .....	45
4.5 Measurement of Twitter censorship .....	46
4.6 Comparative analysis of three social network sites.....	48
4.6.1 Comparative analysis of blocked keywords.....	48
4.6.2 Comparative analysis of types of blocked keywords .....	48
Chapter 5: Discussion and conclusion .....	51
5.1 Discussion.....	51
5.2 Conclusion .....	53
5.3 Future enhancement .....	53
References.....	55

## List of figures

<i>Figure 1: Percentage of blocked porn sites on Facebook</i> .....	40
<i>Figure 2: Frequency distribution of blocked keywords on five country-specific of Google</i> .....	41
<i>Figure 3: Frequency distribution of blocked keywords on Imgur</i> .....	45
<i>Figure 4: Frequency distribution of blocked keywords on Twitter</i> .....	47
<i>Figure 5: Comparison of frequency distribution between Imgur, Reddit and Twitter</i> .....	48
<i>Figure 6: Comparison of frequency distribution between Imgur, Reddit and Twitter on the type of keywords</i> .....	49

## List of Tables

<i>Table 1: List of blocked porn sites on Facebook</i> .....	39
<i>Table 2: List of blocked sensitive keywords in Google Korea</i> .....	41
<i>Table 3: List of blocked keywords on Imgur with categories</i> .....	42
<i>Table 4: Statistics of blocked keywords on Imgur</i> .....	44
<i>Table 5: Statistics of blocked keywords on Reddit</i> .....	46
<i>Table 6: Statistics of blocked keywords on Twitter</i> .....	46
<i>Table 7: Comparative analysis of types of blocked keywords</i> .....	48
<i>Table 8: Statistical results across five networks based on countries</i> .....	50
<i>Table 9: Statistical results across five networks based on content categories</i> .....	50