New Data-driven, Signaling-based Approaches to Social Media Analytics

Benjamin James Lucas (B.Bus, B.Bus, Hons 1st Class).

Submitted in Completion of PhD Management (Marketing).

January 2016.
Statement of Originality

The thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to the final version of my thesis being made available worldwide when deposited in the University’s Digital Repository**, subject to the provisions of the Copyright Act 1968.

**Unless an Embargo has been approved for a determined period.

Signed:
Acknowledgements

I would like to thank my family and friends for their support during my PhD. Special thanks go to my brother Daniel, and my close academic colleagues (and friends) at UoN (in alphabetical order); John Anthony, Jonathon Deans, Natalie de Vries, Brendan Elliot, Gareth Hurst and Ben Kozary. I would also like to thank all the talented researchers I have met, and friends I have made during my two PhD visits to Maastricht University. I hope to continue to strive for your incredible standards and brilliant creativity. I would also like to thank each of my supervisors;

- Thanks to Jamie for guiding through my entire academic research journey and for always believing in and investing in my ideas, and helping them become a reality. Thanks also for always teaching me (and enforcing) the very highest standards of work; this will undoubtedly benefit me for my entire academic career.

- Thanks to Siggi for always providing detailed and honest feedback to help me to develop my research, and for creating new opportunities for me to take my research to another level.

- Thanks to Martin for sharing new perspectives and cutting-edge research approaches, and for investing in my ideas and helping to guide the project to world-class standards.

I would also like to thank a number of people for their help with specific aspects of this thesis. I would like to thank Shannon Hochkins for his assistance in processing the data used in Chapter 2, and for programming the proprietary API interfacing software used in Chapter 3. I would also like to thank Brendan Elliot for his assistance in processing the data used in Chapter. A special thanks also goes to Peter Danaher at Monash University, Josef Šlerka at Socialbakers, Kilian Thiel at KNIME and Dominik Mahr at Maastricht University for their constructive feedback on Chapter 2. Finally, I also gratefully acknowledge Jos Lemmink and members and guests of the Business Intelligence and Smart Services (BISS) Institute for their comments on earlier versions of Chapters 2,3 and 4.
# Table of Contents

## Contents

Acknowledgements............................................................................................................................................................................. 3
Table of Contents.............................................................................................................................................................................. 4
Thesis Synopsis.................................................................................................................................................................................. 1
Chapter 1 – Introduction................................................................................................................................................................... 2
  1.1. Research Problem................................................................................................................................................................... 2
  1.2. Research Objectives and Contributions .............................................................................................................................. 3
    1.2.1. Chapter 2 – Discrete Structural Indexing of Substitutive Language for Social Media Analytics: A Graph-based Text Mining Approach .......................................................... 4
    1.2.2. Chapter 3 - A Picture is Worth a Thousand Nodes: Graph-indexing Terms Identified within Visual Brand Content on Social Media.................................................................................. 5
    1.2.3 Chapter 4 - Analyzing Service Brand-content and Customer Feedback on Social Media with Symbolic Regression: A Substitutive Language Perspective ................................................. 6
  1.3. Thesis Outline......................................................................................................................................................................... 8

Chapter 2: Discrete Structural Indexing of Substitutive Language for Social Media Analytics: A Graph-based Text Mining Approach................................................................................................................... 9
  2.1. Abstract.................................................................................................................................................................................... 9
  2.2. Introduction............................................................................................................................................................................ 10
  2.3. Literature............................................................................................................................................................................... 14
  2.4. Contribution and Requirement Specifications .................................................................................................................. 16
  2.5. Proposed Procedure – Method and Implementations.................................................................................................... 17
    2.5.1. Feature Development and Document Scoring ............................................................................................................. 17
    2.5.2. Combined Unsupervised–Supervised Machine Learning Implementation ............................................................... 19
    2.5.3. Multiplex Network Graph Implementation ................................................................................................................. 20
    2.5.4. Word Co-Occurrence Network Graph Implementation .......................................................................................... 21
    2.5.5. Topology Measure Implementation ........................................................................................................................... 22
    2.5.6. Community Detection Implementation Using Modularity ......................................................................................... 23
  2.6. Illustrative Application of the Proposed Procedure ..................................................................................................... 24
    2.6.1. Dataset ............................................................................................................................................................................ 24
    2.6.2. Results: Combined Unsupervised–Supervised Machine Learning Implementation ......................................................... 24
    2.6.3. Results: Multiplex Network Graphs Application ............................................................................................ 27
    2.6.4. Results: Word Co-Occurrence Network Graphs Implementation ............................................................................. 30
  2.7. Discussion and Conclusions ................................................................................................................................................... 37
    2.7.1. Implications for Marketing Research .......................................................................................................................... 37
Thesis Synopsis

This thesis explores the construction of brand and consumer signals on social media as they exist in different content forms (e.g. textual, visual). Signaling is explored with reference to social identity theory, and the mechanisms by which signals are created are explored with reference to psycholinguistics and semiotics. Central to this pursuit, this thesis employs cutting-edge data-driven methods to analyse the signaling process as it exists in social media. Specifically, in Study 1 this thesis details new combinations and applications of unsupervised machine learning and graph-based text mining approaches to index and rank terms and brands with reference to signaling. In Study 2 this thesis applies existing image recognition technology to the problem of extracting and indexing relevant and salient themes from visual data (images and video) with reference to signaling. In Study 3, this thesis explores alternative approaches to modelling signaling, using data-driven techniques.

Thus, the overarching objective of this thesis is to make contributions to the theoretical development of signaling, with particular reference to the interface between consumer and brand side signaling. Additionally, this thesis contributes to understanding how alternative theoretical lenses can be employed to measure and understand how signals manifest (on both the consumer and brand sides). Secondary to this, this thesis aims to introduce new ways of thinking about how consumer behavior processes are measured, and their manifestations and effects quantified. This is done via the implementation of data-driven analytics techniques to achieve practical outcomes focused on business analytics, business intelligence and business dashboards in a social media context.