Acta Universitatis
Lappeenrantaensis
811

Arash Hajikhani

UNDERSTANDING AND LEVERAGING THE SOCIAL NETWORK SERVICES IN INNOVATION ECOSYSTEMS

Arash Hajikhani

UNDERSTANDING AND LEVERAGING THE SOCIAL NETWORK SERVICES IN INNOVATION ECOSYSTEMS

Dissertation for the degree of Doctor of Science (Technology) to be presented with due permission for public examination and criticism in the lecture room 2303 at Lappeenranta University of Technology, Lappeenranta, Finland on the 5th of October, 2018, at noon.

Acta Universitatis
Lappeenrantaensis 811

Supervisors Professor Helinä Melkas
LUT School of Engineering Science
Lappeenranta University of Technology
Finland

Professor Jari Porras
LUT School of Engineering Science
Lappeenranta University of Technology
Finland

Reviewers Professor Marko Seppänen
Industrial and Information Management
Tampere University of Technology
Finland

Associate Professor Rahul Basole
School of Interactive Computing
Georgia Institute of Technology
USA

Opponent Professor Pekka Kess
Industrial Engineering and Management
University of Oulu
Finland

ISBN 978-952-335-265-0
ISBN 978-952-335-266-7 (PDF)
ISSN-L 1456-4491
ISSN 1456-4491

Lappeenrannan teknillinen yliopisto
Yliopistopaino 2018

ABSTRACT

Arash Hajikhani
Understanding and leveraging the social network services in innovation ecosystems
Lappeenranta 2018
100 pages

Acta Universitatis Lappeenrantaensis 811
Diss. Lappeenranta University of Technology

ISBN 978-952-335-265-0, ISBN 978-952-335-266-7 (PDF)
ISSN-L 1456-4491, ISSN 1456-4491

In today's knowledge-based economies, it is generally accepted that innovations are integral to the foundation of both regional and national economic development, as well as one of the main causes for social and technical transitions. In an effort to boost and benchmark innovation, metrics and indicators have been designed to measure its various stages of development in order to gain insight into what is driving results. In an effort to make such measurements, a systems approach had been adopted in order to capture the dynamic and complex nature of innovation. However, an ecosystem approach has recently begun to attract attention as a framework for studying innovation.

The term "innovation ecosystem" is often employed to explain a large and diverse set of participants and resources essential to the success of any innovation. Literature on innovation ecosystems emphasizes both the importance of a network of linkages between multiple actors and taking a holistic approach to include all players in the ecosystem. This is done to provide synergy, which has an effect on the overall outcome. This dissertation advances the existing research on innovation ecosystems by incorporating the soft aspects of innovation and studying social network services (SNSs) as a complementarity within said ecosystem. SNS platforms (e.g. Twitter, Facebook) provide opportunities for mass communication and interaction, both of which mediate societal discussion. These platforms create a unique opportunity to inform a holistic approach to innovation.

The purpose of this thesis is to discuss the importance of SNSs in innovation ecosystems and attempt to operationalize the valuable data within SNSs for a deeper understanding of innovation. First, this thesis introduces the measurement and evaluation practices used, with particular effort made to highlight how the term "ecosystem" first emerged and then became associated with studies on innovation. To that end, an in-depth analysis of the innovation ecosystem research and citation network was conducted to assess the growing body of literature on this topic. Secondly, this study utilizes SNS data at both the micro- and the meso-level, meaning the company-, community-, and national-level, and provides novel insights. To do so, advanced textual analyses were performed and machine learning models were employed to explore the content of SNSs. These analysis resulted in several

interesting findings regarding the role of content producer and content quality in the overall interaction within SNSs. This attempt to leverage SNSs for data was then furthered to include the design of a metric used to evaluate and establish benchmarks for counties based on entrepreneurial-oriented activity. For a more exploratory approach, SNSs data was analyzed to ascertain whether patterns existed within discussion topics and in proximity over time. Finally, the theoretical impact and methodological contributions to the literature on innovation ecosystems is included to show a novel approach to the use of SNS data. The findings should help scientists and practitioners to engage with SNSs in a more confident manner when an ecosystem-oriented approach is taken to evaluate innovation.

Keywords: innovation measurement, intangible innovation, innovation ecosystem, social network services, text analysis, natural language processing

ACKNOWLEDGEMENTS

This thesis is a reflection of my journey that largely facilitated by the infrastructure that Finland and Lappeenranta University of Technology provided. I am so grateful to have had many people support along the way otherwise this journey would have not been possible.

Firstly, I would like to express my gratitude to my advisors Prof. Jari Porras and Prof. Helinä Melkas for the continuous support of my doctorate study, for encouraging my research and for helping me to grow as an independent researcher. I would also like to thank the rest of my thesis committee Prof. Marko Seppänen and Prof. Rahul Basole for their insightful comments and encouragement, but also for the hard questions which incented me to widen my research from various perspectives. Special thanks to Prof. Pekka Kess for acting as an opponent in my public defense examination.

My sincere thanks also goes to Prof. Kirsimarja Blomqvist, Prof. Paavo Ritala, Prof. Anne Jalkala, Prof. Laura.Albareda, Dr. Jukka-Pekka Bergman and Dr. Arho Suominen whom I had a chance to collaborate and enjoy the inspirational discussions and support which extended my research perspectives.

I have had the privilege of spending six months as a visiting scholar at the Scandinavian Consortium for Organizational Research (SCANCOR) at Stanford University. I owe a special thank you to many individuals; Anni-Kaisa Kähkönen and Anna Brattström thanks for your time and discussions about my research. Thanks Maude Engström, Mitchell L. Stevens, Sarah Soule and all other SCANCOR scholars for inspiring discussions that challenged me and broaden my horizons. Special thanks to Kristiina Herold for being such a good and supportive friend and a mentor. Dr. Martha Russel thanks for engaging me in the exciting activities in MediaX at Stanford, I am appreciative of our successful collaboration during my visit and I am hopeful about the extended outcome of this collaboration too.

My doctoral educational process was enriched with various short research visits and meeting wonderful people. I would like to thank Research Infrastructure for Research and Innovation Policy Studies (RISIS) EU project and the tremendous community and educational events that it provided among which, Department of Policy Analysis and Public Management Bocconi University, Department of Management, Economics and Industrial Engineering of Politecnico di Milano, CNR IRCRES-Research institute on Sustainable Economic Growth (Rome) and Computer Science Department Vrije Universiteit Amsterdam. I am also grateful to meet Prof. Bob Walrave at Eindhoven University of Technology whom provided a great intuition and insight in the very early stages of my research.

I gratefully acknowledge the financial support, which I have received from the Foundation for Economic Education and LUT Tukisäätiö. I am also thankful to the Erasmus+ for funding my research visits in Europe.

Prof. Juha Väätänen, Prof. Tuomo Kässi, Dr. Daria Podmetina and Riitta Salminen, my LUT journey started in 2011 in GMIT master programme by meeting you. I am thankful for the motivation and support you provided otherwise this would have not been accomplished.

I am also lucky to have been surrounded by such a great colleagues at the LUT. The multidisciplinary nature of my research gave me the chance to get to know and work with people at various schools. It's a great privilege to work with people who you can also call your friends; Antti Knutas, Juho Salminen, Jouni Ikonen, Maria Silva, Timo Hynninen, Jussi Kasurinen, Thank you!. I want to also thank the other staff at the University, especially Tarja Nikkinen, Saara Merritt and Sari Damsten for providing unfailing support for all the day-to-day activities as well as the dissertation process. Special Thanks to Kimberly Rehak for helping in polishing the dissertation.

I have been honored to have great friends and fellow doctoral students; Rahul, Arun, Amrita, Bardia, Pani, Mohammad, Hanieh, Lili, Saeed, Sina thanks for being such wonderful people. A big thank you also to Mikko, Ashok, Päivi, Jenni, Niko, Joona, Jystina, Ekaterina, Pekka, Argyro, Anikka, Janne, Timo, Pontus, Elina, Tamara, Seba, Coni, Shqipe thank you for being there and providing support as well as joyful moments outside of working life.

I want to thank especially my parents and brother, who have always been my backbone during times of need. And most importantly, my greatest gratitude goes to the dearest person in my life, Samira. You are a truly amazing person, you carried me through the everyday ups and downs and lived through the whole process with me. This has been very intense for both of us—thank you for staying by my side.

Arash Hajikhani
September 2018
Helsinki, Finland

No problem can be solved from the same level of consciousness that created it.

Albert Einstein

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PUBLICATIONS

This thesis is based on the following papers. The publishers have granted the rights of these papers to this dissertation.

I. Hajikhani, A. (2017). Emergence and Dissemination of Ecosystem Concept in Innovation Studies: A Systematic Literature Review Study. Presented at the Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS) 2017.

II. Hajikhani, A., Suominen, A. (2018). General System Theory Attributes in Innovation Ecosystem Research Landscape: A Bibliometric and Content Analysis of the Literature. *Submitted to Journal of Technological Forecasting and Social Change*

III. Hajikhani, A., Porras, J., & Melkas, H. (2017). Brand Analysis in Social Network Services: Results from Content Analysis in Twitter Regarding the US Smartphone Market. International Journal of Innovation and Technology Management: Vol 14, Issue 02. doi: 10.1142/S0219877017400089

IV. Hajikhani, A. (2017). Efficiency Assessment of the Social Capital Capacity on Entrepreneurial Activity: A Perspective Driven from Social Media. 2017 Portland International Conference on Management of Engineering and Technology (PICMET). Portland, OR, 2017, pp. 1-10. doi: 10.23919/PICMET.2017.8125317

V. Hajikhani, A., Silva, M. & Porras, J. (2018). Crowd Intelligence Participation in Digital Ecosystem: Systematic Process for Driving Insight from Social Network Services Data. Presented at the Academy of Management Specialized Conference on Big Data and Managing in a Digital Economy.

In this thesis, the publications are referred to as Publication I, Publication II, Publication III, Publication IV and Publication V.

Author's contribution

Arash Hajikhani was the first author of all papers. As principal author and main investigator, I performed all stages of the research process, including: planning and executing data collection, data analysis and the writing process. In publication II, the revisions were handled with the help of a co-author. Likewise, the co-authors of publications III and V were involved in the revision process of these papers.

PART I: OVERVIEW OF THE DISSERTATION

1 INTRODUCTION

This introductory section provides the theoretical background and the research gaps that this dissertation will address. Then, the research objectives and research questions are stated and the conceptual background of the thesis is articulated. Lastly, the overall outline of the thesis will be presented.

1.1 Research background

Since the advent of "the knowledge-based economy," the challenge of measuring and evaluating this economic concept has been core to the discussion (Skolnikoff, 1993; OECD, 1996; Foray, 2004; Leydesdorff, Dolfsma and Van Der Panne, 2006; Shapira *et al.*, 2006). Considered one of the major drivers of economic growth, innovation has gained recognition for its determinants and role in structural change (OECD, 2010b). Both national and international agencies have designed indices in an effort to evaluate innovation. Their reports are comprised of surveys and in-depth interviews partnerships within various organizations and institutions in order to get measurable data (OECD, 2010b). These assessments come in the form of an annual report that ranks countries based on their capacity for and success with innovation or other related concepts. For example, an attempt by the Organization for Economic Co-operation and Development (OECD) to measure innovation performance required various indicators that demonstrated a country's innovative performance. The technology achievement index by the United Nations Development Programme (2002) performed a comparative study of innovation in nations using macro-composites as indicators for the creation of technology. Among the many measurement indices, research and development expenditures, citation-weighted patents, and new product announcements were used in these econometric models and evaluations (Griliches, 1990). Meanwhile, approaches to measuring innovation vary depending on the objective of the evaluation scheme and the author's perceived definition of innovation. For example, the European Innovation Scoreboard (EIS), previously the Innovation Union Scoreboard, employs 20 indicators to evaluate innovation, which provide for a comparative analysis of innovation performance in EU countries (European Commission, 2004). The evaluation indices used by the EIS are comprised of four groups: human resources; the creation of new knowledge; the transmission and application of knowledge; and innovation, finance, output, and markets (European Commission, 2004).

The concept of innovation has been visited as an evolutionary and path-dependent process and, thus, considered complex (Jensen *et al.*, 2007a; Tidd J, 2007; Drucker, 2014). Therefore, an evolutionary approach to economic development and a systems view of innovation have contributed to the concept of complexity. Complexity is characteristically a system or arrangement which features a large number of interacting components (Arthur, 1999). A practical illustration of a systems approach to innovation studies led to the configuration of "national innovation systems," which emphasizes the interactions of a community of actors and institutions that influence the innovative

performance of firms and economies (Lundvall, 1992; Patel and Pavitt, 1994). The recent emerging concept of an "innovation ecosystem" is the result of a paradigm shift that classified innovation as a complex and nonlinear process. The innovation ecosystem was proposed by scholars and practitioners to capture the multiplicity and complexity of the innovation process (see, e.g., Adner, 2006; Iansiti and Levien, 2004; Moore, 1993). This concept reflects the demand imposed by the emergence of knowledge-based economies, in which innovation and its developmental process became increasingly non-linear and network-based (UNDP, 2015). Efforts to study its knowledge structure using complex systems and networks of relations has been done using models such as the university-industry-government relationship, known as the triple helix (Etzkowitz and Leydesdorff, 2000). The triple helix relationship can be formed by the interaction between three (or more) sub-dynamics of a system (Leydesdorff and Etzkowitz, 1996; Etzkowitz and Leydesdorff, 2000). Based on the triple helix model of Etzkowitz and Leydesdorff (2000), Carayannis and Campbell (2009) conceptualized a fourth helix, which emphasizes the role of societal and public discourse, like that transmitted through media and culture, in innovation and the knowledge model. The importance of treating intangibles as capital investments in a national context has been studied extensively by Corrado et al (2006) and Belhocine (2008). Evidently, as much as the tangible and monetary resources are for the accessibility of technology, intangible resources, such as social capital, or simply put the "bonds between people," are of major importance to the innovation process (Claridge, 2004).

Regarding the intangible components of innovation, collective online communication channels, known colloquially as social media outlets, have yet to be leveraged in an ecosystem-oriented approach to innovation studies. Today, social media platforms, such as Facebook and Twitter, are driving new forms of dialogue, social interaction, exchanges of ideas, and collaboration, and, as a result, a large amount of research data. Social media platforms, also called social networking services (SNS), are web-based services that allow users to interact and discuss specific topics, exchange personal information, and share what occurs in their daily lives (Mangold and Faulds, 2009). They assist individuals in constructing a public or semi-public profile within a system and acquire social media friends with whom they share a connection (Boyd and Ellison, 2007). Because these functions can be performed via social media platforms, they provide a rich environment where intangibles, like social capital, can be incubated and cultivated (Burke, Kraut and Marlow, 2011). Social media platforms provide a great opportunity to access mass data. Hence, in the early 20th century, prior to the availability of such data sources, sociologists interviewed people to understand their social connections and, in doing so, formed small social networks for analysis (Pentland, 2014). Today, due to the activity on social networking platforms such as Twitter, it is possible to do a study of SNSs in real time thanks to the exceptionally large amount of content and the millions of nodes and billions of edges available (Kireyev, Palen and Anderson, 2009; Bollen, Mao and Zeng, 2011). In the past decade, thanks to both the rise of computational power and the use of machine learning algorithms to automatized approaches to data analysis, new opportunities for accurate analyses have come about. Concurrently, an exponential growth in internet usage has caused the amount of data to increase exponentially.

Contextually, this thesis focuses on three interconnected concepts based on the literature: measurement approaches to innovation studies; an ecosystem-oriented approach to innovation studies confined to the business and management disciplines; and social media, also called social networking services (SNS), an emerging domain of research motivated by the advent of social media platforms. The focus of this thesis will be on the innovation ecosystem literature and the intangible components which can be materialized by utilizing SNS data.

This dissertation contributes to the growing body of literature centered around innovation ecosystems by expanding on the knowledge of components or complementarities. The focus is on the use of intangible data, such as social network services, as a major component of the ecosystem. The aim is to employ methods that utilize said data for a better comprehension of the innovation ecosystem.

1.2 Research gap

Various disciplines have adopted a systems approach, and it is frequently used in business and management literature—mainly due to its capability to tackle complex problems (Chen, 1975; Goodman, 2015). Systems approaches offer an interdisciplinary way to explore systems in nature, in society, and in many scientific domains as well as a framework with which phenomena can be investigated using a holistic approach (Capra, 1997). In general, the systems approach advocates thinking about a system or an arrangement of related units as a whole rather than focusing on its individual parts. Bertalanffy (1972) made some of the first attempts to discuss applying a systems approach in the social sciences with his General Systems Theory (GST). To comprehensively understand a concept such as innovation, a systems approach has been undertaken to recognize the components of a concept and study them for better comprehension and greater accuracy when evaluating and measuring. The holistic approach toward innovation studies is widely emphasized in the national innovation systems literature (e.g. (Lundvall, 1992; Nelson, 1992; Patel and Pavitt, 1994; Edquist, 1997; Freeman and Luc, 1997)), where multiple external factors influence technological advancement in industry. The ever-increasing number of entities and connections resulting from innovation activities is reflected in new terminology, known as the "innovation ecosystem" (Moore, 1993; Suominen, Seppänen and Dedehayir, 2016; Hajikhani, 2017; Ritala and Almpanopoulou, 2017; Smorodinskaya, Russell and Katukov, 2017). "Innovation ecosystems" have become a trending term within the context of Business Management. The terminology first appeared in policy and business discussions, then and academics followed by conceptualizing and conducting case studies to distinguish and elaborate on this phenomenon (Oh *et al.*, 2016; Suominen, Seppänen and Dedehayir, 2016; Hajikhani, 2017; Ritala and Almpanopoulou, 2017; Smorodinskaya, Russell and Katukov, 2017). The field of business and management adopted the "ecosystem" part of this concept from biology in order to emphasize the interaction of community and interdependent actors, as is found in innovation and entrepreneurship (Autio and Llewellyn, 2014; Khandekar and Phani, 2017). Investigating the possible origins of the use of the "ecosystem" term, several factors can be identified to explain its emergence and adoption in this field.

Innovation is considered a distributed and collective process which involves a variety of components and interaction between them (Freeman and Luc, 1997). Therefore, the term "innovation ecosystem" was proposed to capture the multiplicity and complex nature of the innovation process (see, e.g., Adner, 2006; Iansiti and Levien, 2004; Moore, 1993). A recent conceptualization of an ecosystem approach to innovation studies comes from Adner (2017) who defines it to "the alignment structure of a multilateral set of partners that need to interact for a focal value proposition to materialize."

The holistic perspective on innovation was provided by the reports (e.g. (OECD, 2010a; Scott and Vincent-Lancrin, 2014)) mentioned previously. Observing the various metrics created and used for the evaluation of such a complex concept as innovation begs the question, of which is the many indicators introduced are the most influential. In discussions of the ecosystem in innovation studies, unlike in the preceding holistic approaches, such as the systems approach, a trend to include new and novel data sources is emerging (Evans and Basole, 2016; de Reuver, Sørensen and Basole, 2017). In the era of mass communication which brought about the emergence of social network services platforms, a great opportunity to access (much) publicly available data that reflects society at large is readily provided. Statista, an online portal for market data, states that in 2017, 71 percent of internet users were social network services users. The increased usage of smartphones and mobile devices worldwide has, indeed, led to high user engagement rates on SNS platforms (Internet Scociety, 2014). The Statista report also estimates 2.77 billion users around the globe will use SNSs in 2019, up from 2.46 billion in 2017 (Statista, 2017).

Social network services have become a dominant source of data for governments, corporations, and academics to study human society. In innovation studies, literature and the practice of measuring its intangible aspects has been discussed under different concepts. The concept of the "quadruple helix," supported by the systems approach to identifying components, relationships and functions in innovation systems, places a special focus on social capacity (Carayannis and Campbell, 2009). The quadruple helix (QH) proposed by Carayannis and Campbell is an extension of the "triple helix" concept of university-industry-government relationships initiated in the 1990s by Etzkowitz and Leydesdorff (2000). The fourth helix of QH puts particular attention on highlighting the importance of human and social capital in fostering innovation. Florida (2012) has also recognized the importance of the capacity embedded in human capital and coined the term "creative class" to signify the driving force for economic development in post-industrial cities in the United States. While it is important to include the capacity of human and social capital into an analysis, it is difficult to capture the influence of society on broad concepts such as innovation. Therefore, one flaw within the current evaluations on innovation and ranking reports is that they fail to seriously consider using new data sources, such as social media outlets, to capture such a dynamic construct.

One research gap focuses on the soft aspects of components of the innovation ecosystem, such as social capital. The advent of mass communication platforms, such as social network services, has created an opportunity to have a better understanding of society as

represented in mediums such as SNSs. Concepts like "ecosystem" offer a holistic approach to analyzing innovation and entrepreneurship by incorporating network analytics methods and visualization practices to simplify the complexity involved in these two concepts (Autio and Llewellyn, 2014; Basole *et al.*, 2015, 2016; Thomas, Sharapov and Autio, 2015; Russell and Smorodinskaya, 2018). Innovation ecosystem literature has developed and provided frameworks for better understanding of the types of complementarities and their genesis as they unfold over time (Dedehayir, Mäkinen and Roland Ortt, 2016; Jacobides, Cennamo and Gawer, 2018). This creates an opportunity to enrich the variety of forms of the complementarities. In this study, we take a closer look at SNSs and the data hosted within them as a newly emerged complementarity within the innovation ecosystem.

A second research gap lies within the computational power currently available to process mass amounts of data generated in social network services. While previous research has leveraged the network structure of social media data types to materialize the concept of social capital (e.g. (Hofer & Aubert 2013; Kaigo 2012; N B Ellison et al. 2006; Arora 2016)), in this thesis, focus is put on the textual content within SNSs to investigate everything from the characteristics of the content and content producers to the interaction received. This creates new possibilities for expanding the knowledge on complementarities and, in turn, provides for a comprehensive understanding of the innovation ecosystem.

1.3 Research objective and questions

The "ecosystem" concept was adopted in an effort to capture various stakeholders, their interaction and, therefore, a competitive environment (Jacobides, Cennamo and Gawer, 2018). Ecosystem formation and design are becoming ever more important to the field, so understanding the role of stakeholders and complementarities is an essential first step. The way that interdependent players in an ecosystem work to create and commercialize an innovative end product has been the focus of current research (e.g., Adner and Kapoor, 2010; Kapoor and Lee, 2013; Adner, 2017). Innovation ecosystems are typified by a system-level goal of value co-creation (Lusch and Nambisan, 2015), explicitly and holistically considering the role of complementary providers in value creation and appropriation (Teece, 1986; Jacobides, Knudsen and Augier, 2006). Jacobides et al. (2018) distinguishes ecosystems by the type of complementarities rather than other organizing economic activities Jacobides et al. (2018) distinguishes ecosystems by types of complementarities. The actors in the ecosystems proposed by Jacobides et al. (2018) are multilateral, yet not totally hierarchical, and also customizable on account of their non-generic complementarities. These non-generic complementarities imply a degree of customization, meaning that ecosystems are unique.

The objective of this thesis is to complement past interest in understanding what ecosystems consist of, by understanding a complementarity such as social network services (SNSs) and exploring SNSs' positioning within the innovation ecosystem. Not only does this study investigate the emerging literature on ecosystems within the business

and management discipline but it also reviews and implements methods that capture and evaluate the position of SNSs in innovation ecosystem literature. Furthermore, an effort to operationalize SNS data will be made by studying the central role of its content and the content's characteristics, such as content type and content producer. Extracting the content from SNSs and analyzing its features was performed through advanced textual analysis, such as natural language processing and topic modelling. Furthermore, interactions in SNSs can be benchmarked by investigating various characteristics of the content.

The aforementioned objective can be divided into two major research questions:

1) *What is the role of SNSs as one of the multitude of components in the innovation ecosystem?* Initially, the emergence and adoption of the "innovation ecosystem" within the social sciences will be explored. Then details of the investigation into the characteristics of innovation ecosystem will be provided.

2) *How can SNSs be utilized and materialize as a component in the innovation ecosystem?* To address this question, advanced methods that pulled data from social network services (SNSs) were adopted to gain a greater understanding of the content within SNSs as well as any valuable insights. The practical nature of this research question requires a review of and application of methodology on SNS data. The SNS content examined will introduce the features of this complementary stakeholder in the innovation ecosystem.

The thesis consists of five publications, each of which address its own research question and all of which contribute to the main research objective. The research objective, corresponding research questions and research publications are listed in Figure 1.

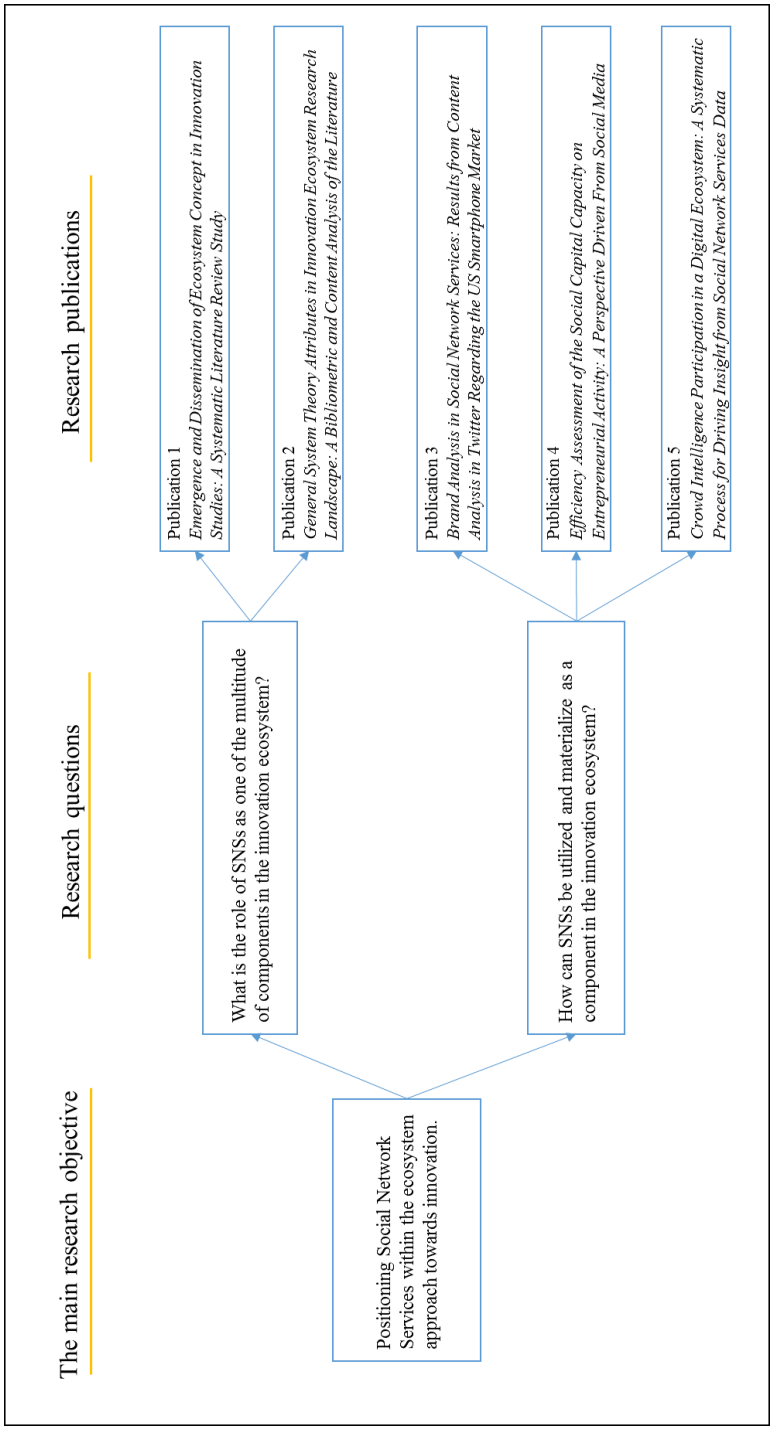


Figure 1. The relationship between the research objective, research questions and publications

1.4 Outline of the thesis

This doctoral dissertation is structured as follows: the thesis is presented into two main parts, the first part gives an overview of the study and second part provides the individual publications; Part I begins with an introduction in Chapter 1, which details the research background, gaps in the research, the main research objective and the research questions of this study; Chapter 2 summarizes the relevant knowledge of the topic and the theoretical background which explores the major research in: innovation measurement and evaluation, the adoption of a systems and ecosystem approach to innovation studies, intangibilities and soft aspects of innovation and the emergence of social network services (SNSs); Chapter 3 discusses the research process and the methodological approaches used as well as elaboration on the data used and the process of data analysis; Chapter 4 summarizes the background and objectives of each individual publication and their consequent contributions to the research; Chapter 5 closes Part I by addressing the research questions and the main contributions of this thesis, the overall research assessment, theoretical and managerial implications, limitations of the study and suggestions for future research. Figure 2 is a visual presentation of the outline of this thesis.

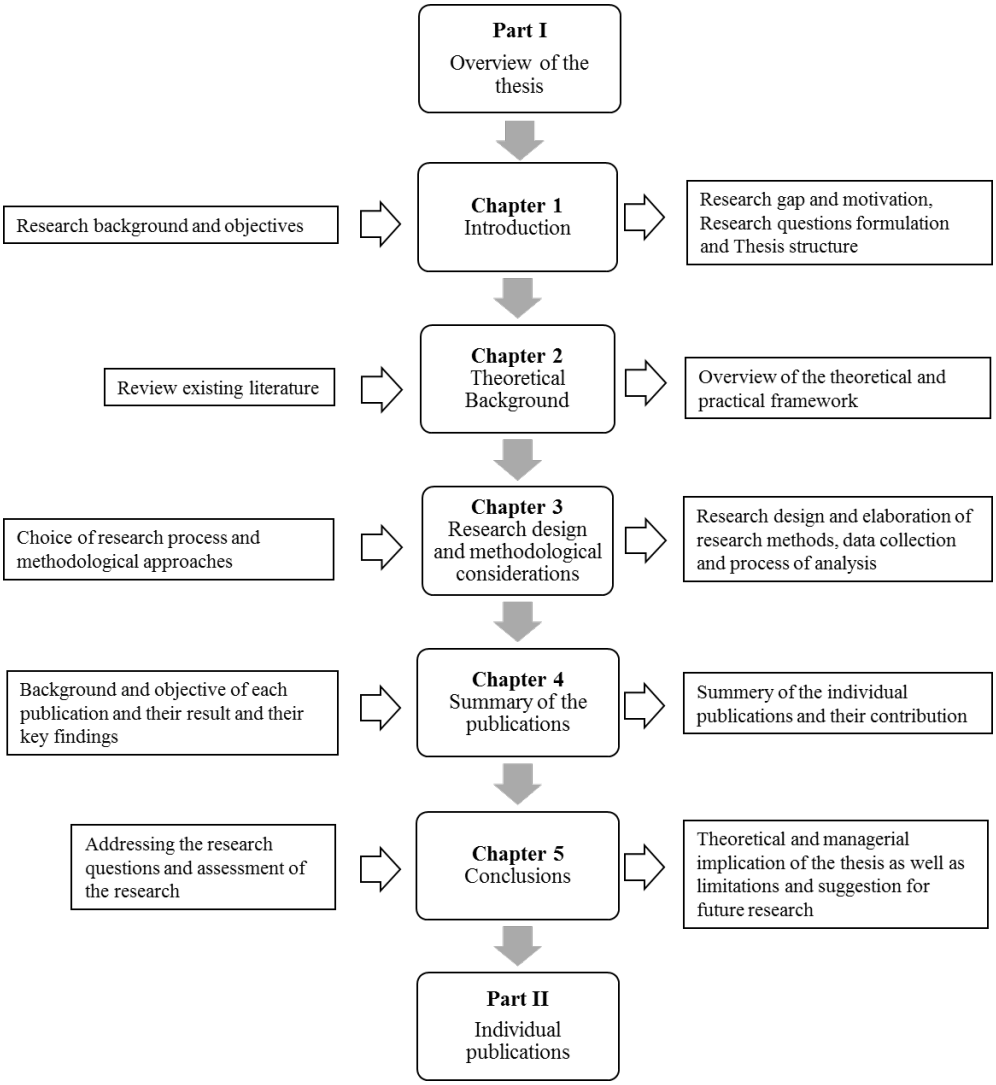


Figure 2. The outline of the thesis

2 THEORETICAL BACKGROUND

In this section, the theoretical background of this thesis will be discussed in detail. Due to the conceptual positioning of the thesis, the discussion is divided into two major sections and their corresponding subtopics, accordingly.

The first section reviews the innovation management literature in the business management discipline and focuses on attempts to evaluate and measure innovation at both the micro- and meso-level. It will then continue by reviewing the challenges met during previous attempts at innovation measurement. Next, both the adoption of the systems approach to studying the complexity of innovation and the later emerging discussion of innovation ecosystems are explored. All in all, an elaboration of the literature suggests the ecosystem approach be utilized when analyzing innovation's characteristics and components. The second section will focus on the emerging entity of social network services (SNSs) and the data hosted by these platforms, which is an underutilized data source and needs to be considered when taking a holistic approach to innovation evaluation. Therefore, an attempt to understand the role of SNSs in capturing an accurate picture of the intangible dimensions of the innovation ecosystem will be made. More specifically, the approach of this thesis is to take the user-generated content in SNSs as a unit of analysis and explore the role of content quality and content producer in the interaction raised by SNSs.

2.1 Innovation measurement and evaluation

The concept of innovation is recognized as a major role player in today's knowledge-based economies (Dahlstrand and Stevenson, 2010). Managing innovation has become a vital task to ensure that organizations endure development, growth and sustainability in today's dynamic environment (Zhao, 2005). Definitions of innovation can vary, but this concept can be perceived as a process of creating value from ideas; on a meso scale, innovation is related to the incorporation of novelties of knowledge in an economy (Tidd and Bessant, 2014). Innovation has been conceptualized in many ways, such as products and services, both organizational or social (Schumpeter, 1934; Koschatzky, 2005; Armbruster *et al.*, 2008; Tidd and Bessant, 2014). Evolutionary economic scholars have discussed that innovation is an evolutionary and path-dependent process (Freeman and Luc, 1997; Jensen *et al.*, 2007).

The attraction of using innovation as a determinant of growth has straightforward empirical effects on its measurement. Innovation is seen as a combination of existing elements; and, therefore various indicators were considered and evaluated in order to measure it due to its broad nature and complexity (Edquist, 1997). The variety of ontological and epistemological positions to investigate, analyze and report on the concept of innovation reveals its complexity and multidimensionality (Wolfe, 1994). More specifically, its complex nature is reflected in the multitude of—often fragmented—approaches to measurement and the evaluation of its performance. It is, indeed, a complex

issue and practice for many modern organizations to quantify, evaluate, and benchmark their innovation competence (Frenkel *et al.*, 2001). The ever-existing challenge is to measure the complex processes that influence innovation's capability for optimal management (Cordero, 1990). Multiple attempts have been made to measure innovation and the practice of administering surveys to do so is among the most popular. A common survey instrument, such as the Community Innovation Survey (CIS) which is based on the so-called Oslo Manual, collected a range of information on the nature and determinants of innovation processes and subsequent performance of firms (Archibugi *et al.*, 1994; Mairesse and Mohnen, 2007). The survey instruments explored firm practices and perspectives, such as: input and output to innovation, corporate strategies, diffusion, and the public policy role in industrial innovation. Innovation surveys offer extensive information on firm practices and attitudes; for example, they demonstrated inter-industries differences in sources of ideas and patterns of innovation (Pavitt, 1984; Klevorick *et al.*, 1995). Measures were constructed by patents and research and development (R&D) expenditures as proxies for innovation performance in manufacturing firms (Lanjouw and Schankerman, 2012). Apart from tangible metrics, Meyer and Harper (2005) endorse a multifactor productivity approach to innovation which emphasizes intangible capital, such as human capital, organizational capital, and intellectual property. Recent attempts to frame innovation measurement practices have also distinguished between tangible and intangible capital (Rose *et al.*, 2009). Due to the evolutionary nature of innovation as a concept, it is difficult to identify a comprehensive body of literature where the discussion of the issues within this practice takes place. An attempt by Milbergs and Vonortas (2004) is among the few found. The researchers attempted to show the development of the indicators of innovation over multiple generations together in one perspective, as can be seen in Figure 3.

1st generation input indicators 1950-1970	2nd generation output indicators 1970-1990	3rd generation innovation indicators 1990-2000	4th generation process indicators 2000-
▪ R&D expenditures ▪ Education/training expenditures ▪ Capital expenditures ▪ Technology intensity ▪ Science and technology substance	▪ Patents ▪ Publications ▪ Scientific and technological activation ▪ Product innovations ▪ High-tech share	▪ Innovation benchmarking ▪ Innovation researches ▪ Primer researches	▪ Knowledge ▪ Intangible assets ▪ Networks ▪ Demand research ▪ Clusters ▪ Management techniques ▪ Risk/return ▪ System approach ▪ Organization culture ▪ Mergers ▪ Wikis

Figure 3. Generational development of innovation indicators. Source: Milbergs (2004)

The evolution of the indicators of development over several generations shows trends towards the acceptance of intangibles as proxies for measuring innovation performance. In the systems concept of processing inputs to outputs to measure innovation, the notion of tangible and intangible constructs was further elaborated. Tangible inputs are characterized as a physical embodiment and cost while intangible inputs are not

represented by a physical embodiment (Blair and Wallman, 2000; Jarboe and Furrow, 2008). Intangible inputs are considered an asset if they engender future benefits (Lev, 2000) and are frequently referred to as "intellectual" assets in business management literature and "knowledge assets" in economic sources (Wilkins, Van Wegen and De Hoog, 1997; Stone, Shipp and Leader, 2008). Jarboe and Ellis (2010) articulated that intangible assets transition to innovation activities in three categories of capital: human, structural, and relational, with the latter representing the knowledge of external stakeholders (collaborators, suppliers, and customers). The added value of external structures along with internal management and organizational process structures as well as individual competences within firms are envisaged in brand assets and relationships with outside stakeholders (Bontis, 2001). A national scale report by OECD (2010a) discusses the struggles countries have measuring intangible investments, such as brand equity, as they are subjected to national studies and do not reflect standardized methods and categorizations.

Adams et al. (2010), has a literature review on the measurement approaches towards innovation, with soft, or intangible, aspects of innovation categorized within the input framework. It is particularly more difficult to find measurement instruments that represent the softer inputs of skills and knowledge. According to Adams et.al. (2006), tacit knowledge input does not seem to be well-recorded using existing measures, and measures of appropriate skill levels have yet to be developed. Therefore, this imbalance created an opportunity to develop a balanced set of measures able to cover all sub-dimensions of the input category (Adams *et al.*, 2006; Ravn, Nielsen and Mejlgaard, 2015). One of the main aims in the thesis is to advance the measurement capabilities of the intangible, or soft, aspects of innovation which have not been materialized so far. The thesis continues the discussion of approaches towards innovation measurement and evaluation with a special interest in capturing influential yet intangible dimensions of innovation.

2.1.1 From a systems to ecosystem approach in innovation studies

Innovation is an increasingly distributed and collective process that involves a variety of components along with interaction between them (Freeman and Luc, 1997). As a result, an increasing number of researchers have started to develop a holistic view of innovation and technological development (Bergek *et al.*, 2008). Accordingly, a wide range of systems configurations have emerged within innovation studies, such as: national systems of innovation (e.g. Edquist (1997), Lundvall (1992) and Nelson (1992)), technological systems approach (e.g. Carlssonand and Stankiewicz (1991)), and the sociotechnical systems approach (e.g. Bijker (1995)) and the network approach (e.g. Håkansson (1990)). Network science has caused an important evolution in innovation studies (Allen, Maguire and McKelvey, 2011) and imposed a shift in focus from manufacturer-centric to network-centric approaches (Powell, Koput and Smith-Doerr, 1996; Iansiti and Levien, 2004). In this recent shift, a "network" is considered a new unit of analysis for studying innovation processes and links traditional actors, such as firms, with new agents, such as users and communities (Andriani, 2011). The network mindset has influenced the systems

approach to innovation where various components that range from people, enterprises and institutions interact in order to turn an idea into a process, product, or service on the market. With a meso-level view of national innovation systems, the agenda of researchers has moved from the level of the individual actor to that of a collective agent, such as: an international (or global), national, regional, or local innovation system. From the perspective of the "national innovation systems" research domain, it has been argued that the technological advancement of industries is influenced by external factors (Lundvall, 1992; Nelson, 1992; Patel and Pavitt, 1994; Edquist, 1997; Freeman and Luc, 1997). In contrast to proxies of innovative activities, like input measures such as R&D expenditures (Mansfield and Griliches, 1984), or innovation outcomes such as patents (Griliches, 1990), innovation systems are now considering many other externalities. The emphasis on innovation being non-linear and network-based, led to the adoption of the ecosystem framework in innovation studies, in an effort to capture the multiplicity and complexity of the innovation process (see, e.g., Adner, 2006; Iansiti and Levien, 2004; Moore, 1993). The term "ecosystem" is used substantially and has lent itself to a wide variety of scientific domains. Indeed, has even been adopted in innovation policy briefings and assessments in the past few years (European Union, 2015, 2016, 2017; Mulas, Minges and Applebaum, 2016; Sworder, 2017). This concept has enriched the approach to innovation systems, which initially placed stress on innovation and technology development in the industrial era (Russell and Smorodinskaya, 2018). National or regional innovation systems used to be seen as static structures measured and regulated by governmental bodies, where successful performance was contingent on mass actor involvement and a deliberate infrastructure (Smorodinskaya, Russell and Katukov, 2017). However, the innovation ecosystem is reflected in dynamic and agile collaborative arrangements that appreciate self-governance as an essential requirement for interactive innovation (Townsend, Pang and Weddle, 2009; Bramwell, Hepburn and Wolfe, 2012; Rinkinen, 2016).

The term "ecosystem" was first applied in the fields of business and economics by Rothschild (1990). In Rothschild's book *Bionomics*, he likens the understanding of economics to that of a biological system. In the same vein, Moore (1993) presented the term "business ecosystem" to highlight the essentiality of competition between components of an ecosystem. Moore further stressed the dynamics that regenerate interactions between organisms and the environment[1]. Based on Suominen (2016), the innovation ecosystem is a distinguished cluster, thriving by itself as a stand-alone domain among clusters, such as the business, knowledge, and digital ecosystems as well as digital platforms (Gomes *et al.*, 2015; Valkokari, 2015; Dedehayir, Mäkinen and Roland Ortt, 2016; Suominen, Seppänen and Dedehayir, 2016). However, the scholarly discourse is not set on this or any other definition for the innovation ecosystem or platform. The biology-inspired concept of the ecosystem suggests connectedness as well as an evolution that emerges due to interaction between key elements in a knowledge system and technology system (Hage, Jerald; Mote, Jonathon E; Jordan, 2013). Ecosystems in an

[1] A systematic co-citation and bibliometric analysis study by Gomes et.al (2015) reveal that most business and innovation ecosystem builds their studies on Moore (1993).

innovation study context have been defined as the collaborative effort of a diverse set of actors jointly working towards the creation of value (Adner, 2017). These actors, whose individual input is connected to create the structure of the innovation ecosystem, range from the focal firm, the customers, and the suppliers to innovators and other agents working as regulators (Moore, 1996). This definition implies that actors and complementarities must cooperate and compete in the innovation ecosystem and that the innovation ecosystem has a lifecycle which follows a co-evolution process (Gomes *et al.*, 2015; Dedehayir, Mäkinen and Roland Ortt, 2016). Jacobides et al. (2018) differentiates between various types of complementarities (unique or supermodular; generic or specific) that affect the existence, emergence and alignment of actors in an ecosystem. Indeed, innovation ecosystems do not necessarily recognize geographical boundaries; and, therefore, can span the globe and integrate a myriad of actors to materialize value (Dedehayir, Mäkinen and Roland Ortt, 2016). One strength of ecosystems and their distinguishing characteristics is the elimination of need for vertical integration within complementarities. Ecosystem structure, after all, provides a space for complementarities in production and/or consumption to be contained and coordinated (Jacobides, Cennamo and Gawer, 2018). Hence, this platform typically assigns consumers a value creating a hierarchical integration. Individual entities in an ecosystem simultaneously assume the role of provider as well as consumer of products and services (Autio and Llewellyn, 2014; Gawer, 2014).

Overall, the adoption of the ecosystem concept by those who study innovation can now allow for a look beyond the technological aspects and include socio-technical regimes and non-technological elements (strategy, cultures, organization, and institution) to show the capability of the innovation ecosystem (Phillips, 2006; Carayannis and Campbell, 2009). Carayannis and Campbell (2009) pinpoints intellectual capital as well as cultural and technological artifacts as essential embodiments of the innovation ecosystem in the era of knowledge-based societies and economies. A comprehensive review by Russell and Smorodinskaya (2018) on categorizing systems and the ecosystem approach to innovation was conducted. Table 1 lists ten different aspects for systems and ecosystem approaches. The major difference between the two categories is the static/linear vs complex/non-linear, which were adopted by the traditional system and the new ecosystem approaches, respectively.

Table 1 Ecosystem vs system approaches to innovation studies. Author edited version adopted from (Russell and Smorodinskaya, 2018)

	Systems approach	**Ecosystem approach**
Economic dynamics	Linear systems - closed, static, in equilibrium	Non-linear systems - open, dynamic, dissipative
Emergence and synergy	Macro-level growth patterns are formed by linear summation of individual decisions of homogenous agents, with few synergies occurring spontaneously	Macro-level growth patterns emerge nonlinearly, out of synergies generated by dynamic network interactions of various heterogeneous agents at the micro-level Network
Predominant model of economic governance and adaptation	Hierarchic model: a rigid, centralized organization governed by administrator through top-down decisions. The economy lacks feedback linkages for self-	Heterarchical model: a dispersed, agile network with spontaneous self-organization, self-regulated through horizontal coordination of network nodes and

	adjustment to changing environment and, hence, has low capacity for adaptation	collaborative consensus-building. The economy gets self-adaptable through interactive communication of agents, their feedback, their learning and proactive reciprocity
Network interactions	Network relationships are inessential, agents interact indirectly through market price mechanisms	Network relationships are essential, economic systems of all levels (from local to global) are seen as network-based ecosystems meant for innovation
Interpretation of innovation	Limited endogenous capacity of economic system, dependent on a complex of its available resources. Requires external incentives or exogenous sources, not connected with a system's social and structural transformations. Implies linear process of knowledge flow, from science to industry ('mode 1' in knowledge creation)	Sustainable endogenous capacity of economic system, based on internal incentives and new sources, arising from a system's ability for continual self-correcting structural changes. Implies non-linear process of knowledge flow ('mode 2'), relying on interactive communication of various agents, as well as continual, systemic process ('mode 3'), arising from proliferation of collaborative networks and their ecosystems
Model of producing innovations (goods, values, technologies) Interpretation	Linear models of innovation ('technology push' and 'demand pull'), driven by technological developments of individual firms	Interactive model: co-creation of innovations by networked agents through their collaboration within a generated ecosystem of linkages and assets
Interpretation of innovation systems (regional, national, macro- regional)	Non-cohesive organizational structures that depend on the involvement of a certain critical mass of agents, talents and new infrastructure	Holistic social communities, or ecosystems, with the properties of complex adaptive systems, depending on a certain critical mass of interactive inter-linkages among networked agents
Institutional and business environment for innovation	Creation of new institutions, technologies and industries is higher priority than enhancement of cohesive context for a smooth dissemination of innovations across sectors and regions	Priority is given to continual improvements in an environment, with the purpose to eliminate barriers and provide incentives for more business networks, more collaboration, more cohesion, and continual knowledge spillovers across and around the economy
Focus of strategies for innovation and growth	To develop R&D and national innovation systems by supporting its agents and infrastructure elements, with no focus on collaboration and its innovation synergy effect	To promote localized ecosystems across the economy and enhance their innovation synergy effects by facilitating the dynamics of interaction and collaboration within/between networks
Data source type associated with studies	Macro-level data, such as yearly reports of countries'/regions' performance based on static design indicators. Survey-based. Data mainly on R&D and patenting activity	Utilizing data sets with characteristics such as network-based, real-time data sets and mass data. Social network services data which deliver micro interactions among entities. Web data-based approach to investigate business ecosystem dynamics

In spite of efforts to adopt the ecosystem approach to study innovation, as it encourages systems thinking and a willingness to learn from a multitude of entities, a lack of rigor and measurement schemes in these studies predominate (Oh *et al.*, 2016). Various attempts have been made to capture the complexity of the innovation ecosystem, such as applications of Social Network Analysis (SNA) and network visualizations (Russell *et al.*, 2011; Evans and Basole, 2016; Huhtamäki, 2016). SNA is the process of investigating a network structure of various entities and their connections and then utilizing that data

to gain a better understanding of ecosystem transformations and the evolution of interaction over time (Kolleck, 2013; Arora, 2016; Huhtamäki and Rubens, 2016a; Xu *et al.*, 2017). While SNA offers various metrics for comprehending the innovation ecosystem, visual representation of ecosystem networks is also important for understanding network data and conveying analysis results (Huhtamäki, 2016; Huhtamäki and Rubens, 2016).

2.1.2 Soft aspects of innovation and intangible data

The evolving process of measuring indicators of innovation has witnessed a shift in focus from tangible indicators, such as scientific publications and patents, to more intangible indicators, like brand presence as well as human and social capital. The system approach towards innovation resulted to a holistic view which involved more components for innovation comprehension and evaluation purposes. The recent discussion under the concept of innovation ecosystem provided the groundwork for the elaboration of the alignment of actors, activities and linkages (Adner, 2017; Jacobides, Cennamo and Gawer, 2018).

In this section, the focus is on the intangible aspects of innovation. Intangible, or soft, aspects of innovation have yet to gain much attention; nonetheless, investigating intangible aspects is crucial for making an accurate evaluation of the innovation process and competitiveness in a knowledge-based economy (den Hertog, Bilderbeek and Maltha, 1997; Canibano *et al.*, 1999). Soft aspects in innovation studies have gained more attention in the past decade (Grimaldi *et al.*, 2017), where a behavioral approach to innovation was taken (Sundbo, 2006), products in creative industries were investigated, (Stoneman, 2010) and an overhaul in innovation policy occurred as a result (NESTA, 2009).

Among the soft aspects of innovation's many dimensions, realizations of innovation outcomes are attributed to humans and society at large (Schuller, 2001; Kaasa, 2007; Wardhani, Acur and Mendibil, 2016). The importance of human capital is due to its ability to decode and adopt new information (Becker, 1974; Jones, 2001; Klein and Cook, 2006). From an economic-growth perspective, human capital and it's ability to process information is crucial to the innovation process (e.g., (Nelson and Phelps, 1966; Romer, 1990; Grossman and Elhanan, 1993)). The absence of quality human capital will result in high costs for innovation (Shane, 2000; Fairlie and Robb, 2007; von Hippel, 2007).

Human capital has been measured and utilized mainly by proxies, such as: level of formal education, number of years in school, or years at a job or current firm (Corbett, 2007). Fundamentally, human capital is the investment in knowledge and skill development that will consequently result in economic benefits for both the individual and the collective (Schuller, 2001). On a different note, social capital considers the network connection between individuals and entities, the relationships between them, and the norms which govern these relationships. Social capital basically attempts to get the value of social relationships and social networks to complement economic capital for economic growth.

The concept of "social capital" is accredited to Putnam (1993) in political science, Coleman (2000) in educational sociology, and Fukuyama (1995) in economic history and sociology. Social capital has been most concisely and measurably defined as "resources embedded in social networks" (Lin, 2002). A definition by Putnam (1993) presents social capital as "features of social organization, such as trust, norms, and networks that can improve the efficiency of society." The role of social capital in the development of an economy emphasizes the diffusion of knowledge, the reduction of transaction costs, and the discouraging of opportunistic behaviors (Durlauf & Fafchamps 2004; Granovetter 1985; Coleman 2000; Putnam 1993; Fukuyama 1995).

In fact, the prevalence of social capital is such that some have added it as a "fourth helix" to the traditional "triple helix" model of knowledge, where "university-industry-government relations" are the foci. The "quadruple helix" model of knowledge stresses the "media-based and culture-based public" as a new helix. Described by Carayannis and Campbell (2009), the fourth helix contains the capacity that culture and its values have and how "public reality" is constructed and communicated by the media—both of which certainly have influence over a national innovation system. Figure 4 is the conceptually-based illustration of the fourth helix into the initial triple helix model of knowledge with university-industry-government interaction[2].

[2] The Fifth Helix (Quintuple) has continued the discussions by emphasizing the natural environment or a social ecology (Carayannis and Campbell, 2009; Carayannis and Economy, 2014).

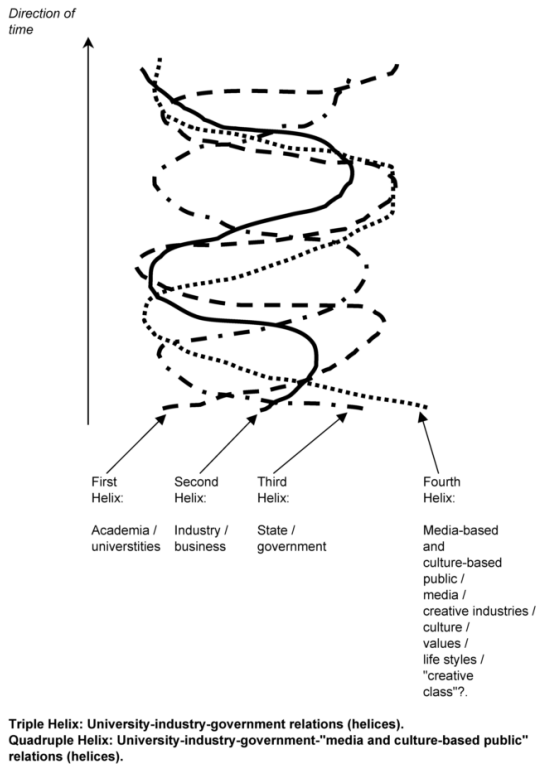


Figure 4. The conceptualization of "Quadruple Helix" adopted from (Carayannis and Campbell, 2009)

Further deliberation of Carayannis and Campbell (2009) associates the fourth helix with "media," "creative industries," "culture," "values," "life styles," "art," and, perhaps, also the notion of the "creative class" (presented by Florida (2012)). Therefore, having an advanced knowledge-based economy would require an appropriate "innovation culture" (Castellacci, 2014). Public discourse transmitted and/or interpreted by media is considered crucial to a society's advocacy for innovation and knowledge (research, technology, education) (Del Giudice, Carayannis and Della Peruta, 2012). Despite the discussion and recognition of intangibles, like social capital, in various studies (including innovation studies), there is still no consensus on its measurement. Various research studies have recognized the measurement of social capital as a central to the measurability dilemma (Schuller, 2001; Claridge, 2004; Stanley and Briscoe, 2010; Hansen and Roll, 2016). Based on its definition, social capital is a dynamic and nonlinear concept that attempts to depict a network of relationships from many components. This definition calls for an extension of the sources of data collection and an incorporation of advanced methods for analyzing this novel data set.

2.2 ICT advancement and emergence of social network services

The rapid development of information and communication technologies (ICT) resulted in one of the biggest changes to humanity (IT for Change, 2013; Archer, Geoff. Van Woensel, Van Woensel and Archer, 2015; Niall, 2016; Marr, 2017). With the immersive invasion of ICT and its applications, a transformation occurred in human communication, giving rise to the new trend of social network service platforms. Mass communication platforms, such as social media or social network services (SNSs), facilitate the creation and sharing of information via virtual communities and networks (Lievrouw and Livingstone, 2002; Pentland, 2014). Extensive participation in SNSs has widened the scope of knowledge sharing and collaboration and created an opportunity to not only affect change but also challenge social norms (Breuer, 2011). Approximately 2 billion people are using one of the major SNSs platforms, and these figures are expected to grow due to a global increase in both internet coverage and mobile device usage (Statista, 2018). Based on the ratio which represents site visits and time spent on a particular platform, Facebook and YouTube are among the top three websites worldwide, with Twitter and LinkedIn in eighth and thirteenth positions, respectively (Alexa, 2018).

Thus, academia has paid much attention to research on social media or SNSs. Taking only a cursory glance at the scientific publications indexed on the Web of Science (WoS), it is easy to see that the number of papers noticeably increased after 2014. A constructed query included “SNSs” and “social media” along with other popular applications as keywords. The query ran on 15.04.2018 on Web of Science resulted in 15,500 records (including articles, proceedings and book material). Figure 5 represents the distribution of volumes published per year.

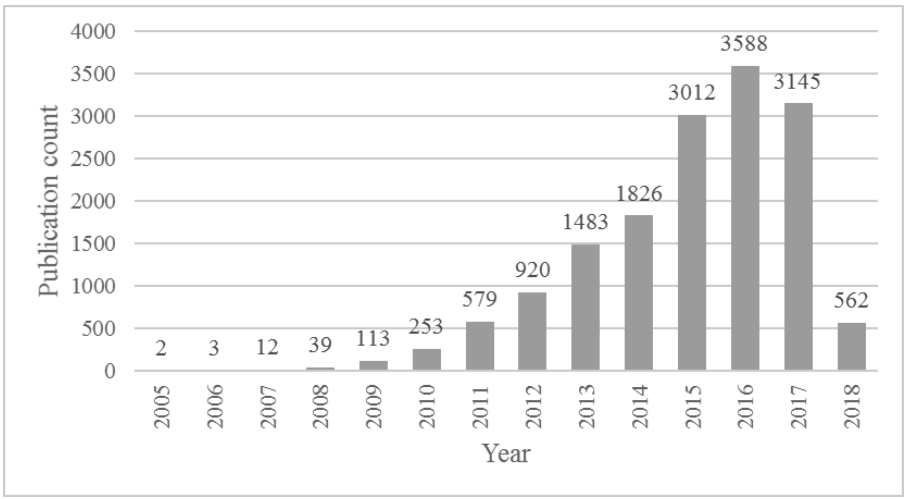


Figure 5. Number of published articles over the years that discuss SNSs-related topics in Web of Science

The exponential growth in publication volume over the years is clearly visible in Figure 5. The upward trend is highly associated with the advent of SNSs tools, which also appeared in the mid-2000s (e.g. Twitter was founded in 2006 and Facebook was founded in 2004). As is shown in the WoS citation report (Figure 6), an analysis of 9,916 articles revealed as many as 41,445 citing articles, indicating a high penetration of the field.

Figure 6. Citation report generated by Web of Science core collection

A closer look at the research areas of the publications and their citing articles revealed that getting a good sense of the distribution of articles by field is possible. Figure 7 shows the distribution of publications and their citing articles. The distribution of fields shows a wide range of research areas. Looking at SNSs-related publications, a ranking of publication volumes indicates communications and computer science as the research areas with the highest number of publications. However, relative to number of publications, areas such as computer science, business economics and engineering are showing a higher number of citations being generated.

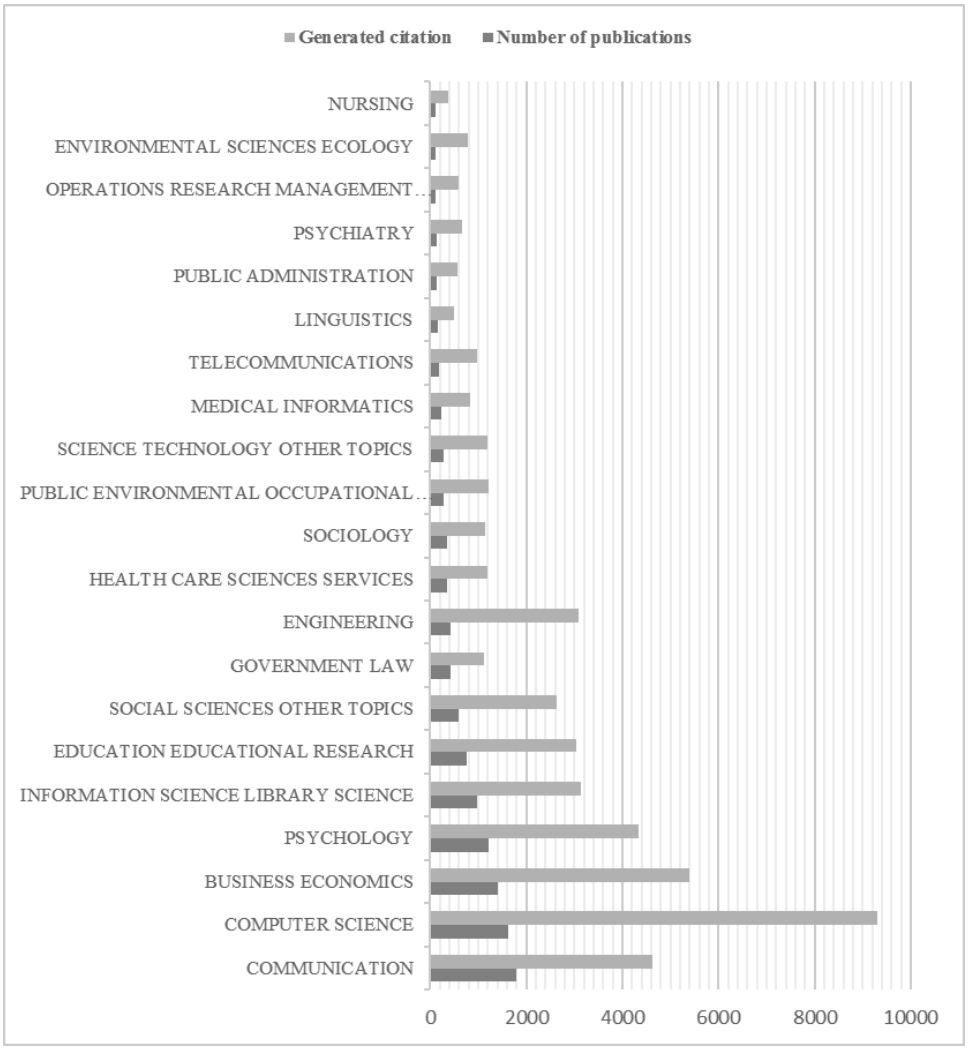


Figure 7. The number of publications and generated citations distributed by research subject

A comprehensive look at social media research done by Ngai et al. (2015) resulted in a better understanding of the theories, constructs and conceptual frameworks that are referenced in the literature. A framework is used to understand the causal relationships among different research constructs adopted so far in the literature. An illustration is provided in Figure 8 to show the positioning of the framework for social media research.

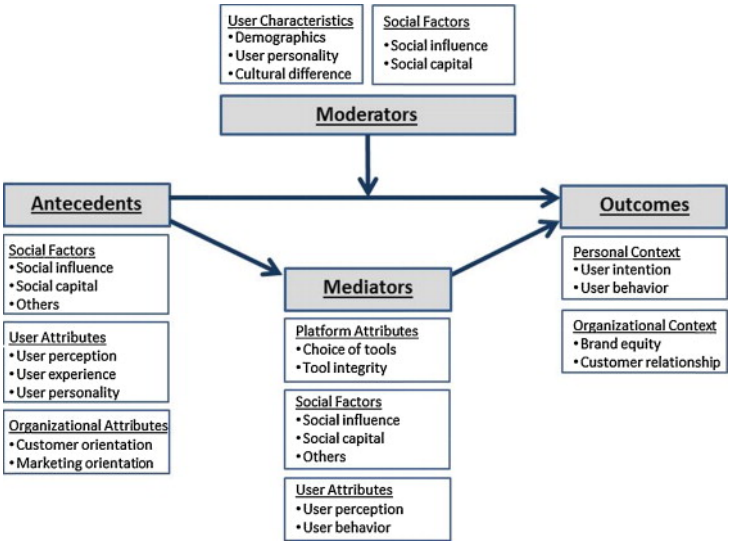


Figure 8. Framework for social media research. Adopted from (Ngai, Tao and Moon, 2015)

Based on the literature review of social media, a causal-chain framework was developed to illustrate the inter-relationships among the research constructs adopted. This study attempted to structure previous research and proposed conceptual models into a comprehensive framework. From Figure 8, this attempt at comprehensiveness can be seen through the categorization of social capital as both an antecedent and a mediator as represented in previous social media research.

The literature has discussed the importance of computer-supported communication and the concept of social capital (Kobayashi, Ikeda and Miyata, 2006; Carmichael, Archibald and Lund, 2015). The network structure of social media communication platforms can have instant social implications, as it provides the opportunity to establish new networks which leads to a so-called bridging or bonding of social capital online (Quan-Haase *et al.*, 2001; Ellison, Steinfield and Lampe, 2006; Williams, 2006; Pénard and Poussing, 2010; Kaigo, 2012; Amichai-Hamburger, Kingsbury and Schneider, 2013). Social capital is broadly defined as the set of resources embedded in the relationships among actors within a network (Robert, Dennis and Ahuja, 2008). A three-dimensional view of social capital categorizes it as composed of structural, relational, and cognitive dimensions (Wasko and Faraj, 2005). The cognitive dimension of social capital is a shared perspective, understanding, representation, interpretation, and system. Examples of this are language, code, goal, culture, vision, expertise, tenure, and narrative. Publicly-available data sources, such as Twitter, have facilitated massive amounts of data collection which can bring the research to the intersection of social sciences, data sciences, and indicator design. As this data is not easily collected through tradition research means (e.g., surveys, interviews, focus groups), it has the potential to shed light on subject areas, such as public opinion and trending topics, that were previously difficult to gather from the population at large (Yi *et al.*, no date; Wiebe *et al.*, 2003; Acosta, 2014; Eichstaedt *et al.*, 2015).

Arguing that social network services should be examined as a potential representation of social capital, this thesis introduces social network services as an essential complementor in an innovation ecosystem.

2.3 Positioning social network services in the innovation ecosystem: Synthesis and a conceptual framework

According to the innovation ecosystem concept, complementarities are integrated across a value chain and can potentially enrich the ecosystem as a whole and determine its success (Adner and Kapoor, 2010). Furthermore, complementary assets relationships can exist in overlapping ecosystems (Kwak, Kim and Park, 2017). Smorodinskaya et al. (2017) went so far to claim that all ecosystem models are complementary and predetermine each other in terms of design, functionality, and pattern of collaboration—a conclusion reached after observing variation in three different formations (platforms, clusters, and value chains). The nature of these complementary components can affect an ecosystem's emergence or growth (Adner and Kapoor, 2016), meaning relationships among participants can be initiated and strengthened as the ecosystem grows as opposed to remaining a hub platform and components. More specifically, the sharing of customers, suppliers, and other stakeholders as well as resources across this network of complementarities can lead to the growth of an innovation ecosystem and value creation.

A balanced development of complementarity value functions is vitally important to the growth of the innovation ecosystem. Therefore, it is crucial to first diagnose the innovation ecosystem complementarities according to an integrated value chain. In this thesis, the process of discovering complementarities showed the importance of Social Network Services (SNSs), as was discussed extensively in the previous section. The availability of mass communication platforms, such as SNSs, makes adjustment and membership; transaction of public opinion, and, therefore, creation of massive amount of content possible. Public communication and the interactions which occur and are hosted on SNS must be considered when an innovation ecosystem is studied. In order to observe such a complementarity as SNSs, the General System Theory (GST) framework was utilized to conduct an in-depth observation of SNSs in various contexts. As elaborated in the previous section, the premise of GST is that complex systems share organizing principles which can be discovered. Based on a review of the GST literature, the five major attributes of a system are categorized for the purposes of this dissertation in table 2 as follows:

Table 2 System Attributes and Definitions According to GST

GST Attributes	Definition
Attribute #1 (Components)	A system consists of interacting elements. It is a set of inter-related and inter-dependent parts arranged in a manner that produces a unified whole. Other names to describe this attribute are: actors, components, users, stakeholders (positions), and entities.
Attribute #2 (Interaction)	The various sub-systems should be studied in their inter-relationships rather than in isolation from each other. The emphasis is on relationships,

	communication, and the process in which the system generates links, activities, and interactions to describe the interacting elements.
Attribute #3 (Dynamism)	A system does not exist in a vacuum. It receives information, material, and energy from other systems as inputs. These inputs undergo a transformation process within a system and leave the system as an output to other systems. This has also been referred as the status of continuous change in the system.
Attribute #4 (Environment)	An organizational system has a boundary that determines which parts are internal and which are external. An organization is a dynamic system as it is responsive to its environment. It is vulnerable to change in its environment. This attribute considers the notion of the boundary/border of the system.
Attribute #5 (Goal)	A desired result or possible outcome that a system envisions to achieve is called the state of a goal in systems literature. This attribute points out the purpose and mission of the system structure.

The five attributes are part of the seminal works on innovation ecosystems conducted by Adner (2006; 2010). Components have been explicitly referred to as such (either complementary or intermediaries) among firms and end customers. Interaction is used in collaborative networks and collaborative arrangements in the context of reducing the cost of firm coordination. Dynamism is discussed as an evolutionary process within the ecosystem that iteratively occurs as a means of assessment to maximize value proposition. Environment is accounted for through focused case studies. Finally, goal is brought into the context of the study (Adner (2006; 2010)) by way of risk assessment and expectation alignment. Figure 9 tries to illustrate the synthesis of complementarities observed in an innovation ecosystem.

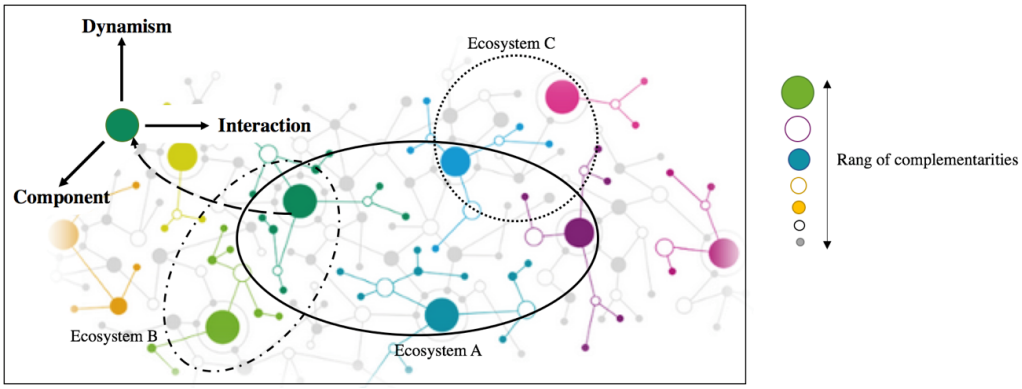


Figure 9. Complementarity Characteristics in Innovation Ecosystem

Figure 9 tries to visually represent the complementarity characteristics taking into consideration the aforementioned GST attributes within an overlapping innovation ecosystem. It shows a range of complementarities and components interacting within one boundary with which other ecosystems might overlap at various levels (firm, economy, society).

An attempt to identify and distinguish the ecosystem approach to innovation studies required looking at bibliometric data and author network collaboration (Publication I). Furthermore, this thesis suggests that GST and its attributes can provide a powerful perspective and methodological lens for the analysis of the emerging ecosystem approach to innovation studies. Therefore, by introducing a system's main attributes, it is possible to detect and examine attributes in the innovation ecosystem literature (Publication II). In the orchestration of various components and complementors, this thesis focuses on observing SNSs and their respective GST attributes. This observation starts by identifying an entity in the innovation ecosystem. Its position with the ecosystem is determined by looking at its sub interactions and the type of dynamism on display. Then the environment or the context and the goal or objective which the innovation ecosystem is set to achieve is observed. The SNSs are studied in three environments in three individual publications (III, IV and V). Table 3 demonstrates the GST attributes for SNSs complementarities in each of the publications.

Table 3 SNSs complementary positioning regarding GST's attributes

System Attributes/Study	Publication III	Publication IV	Publication V
Attribute 1: Components	Social Network Services data (Twitter)	Social Network Services data (Twitter)	Social Network Services data (Twitter)
Attribute 2: Interaction	Sub components in SNS such as users, interaction through communication, Users interacting with each other's content	Sub components of SNS, various actors having discussions on entrepreneurially-oriented activities	Sub components in SNS, various actors having discussions on disaster incidents
Attribute 3: Dynamism	Public availability of SNS data. Live stream nature of SNS data	Public availability of SNS data. Live stream nature of SNS data	Public availability of SNS data. Live stream nature of SNSs data
Attribute 4: Environment	On the context of the study, communication on 5 flagship products of major mobile phone manufacturers on Twitter	Entrepreneurially-oriented discussions on Twitter, for example: of European countries.	Communication on Twitter about a disastrous incident.
Attribute 5: Goal	Firm-level	Economy-level	Societal-level

For the purposes of this study, attempts to materialize previously-discussed features embedded in SNSs as representative data sources for intangibilities, such as social capital, are made. Furthermore, the methodological approach of this thesis leverages SNSs' textual content in order to conceptualize the cognitive dimension of social capital.

3 RESEARCH DESIGN AND METHODOLOGICAL CONSIDERATIONS

The purpose of this chapter is to set out the research design and methodology used in this research. With a variety of reference books purporting various methodologies, it quickly became evident that there is no straightforward way to carry out a research project. Hence, it is important to be aware of the choices made, approaches taken, and methods selected, the impact of which needs to be able to justify said selection. In this research study, Design Science Research (DSR) was selected as the philosophical approach in order to discover and/or recognize opportunities and difficulties relevant to the conceptualization and characteristics of an ecosystem approach to innovation studies. The problem and, therefore, motivation of taking holistic approach to innovation establishes a link to the theoretical explanation and practical utilization of social network services data as a component for evaluation.

DSR is a research methodology in which a designer creates a novel construction to tackle problems faced in the real world, thereby contributing new knowledge to the discipline in which it is applied (Lukka, 2003; Hevner, A.; Chatterjee, 2010). The philosophical assumption of the DSR paradigm implies problem solving and engineering. Simon (1996) and Peffers et al. (2006; 2007) emphasize the problem-solving and engineering nature of DSR, and March and Smith (1995) considered the DSR approach a way to explore for alternative solutions to solve problems that is not only relevant but also improves the problem-solving process. The application of DSR is most notable in the information systems and engineering disciplines, though is not restricted to these fields and is, indeed, found in many others (Vaishnavi, Kuechler and Petter, 2005; Kuechler and Vaishnavi, 2008). DSR has also been shown to develop support for organizational problem solving and has become increasingly relevant to research on products in academic management (Van Aken, 2005; Anne, David and Joan Ernst, 2006; Avenier, 2010).

There are certain rules set by Hevner (2004) for doing DSR which first emphasize the instantiation of an artifact to address a problem. The artifact needs to be novel and relevant to the solution and should undergo a development phase that is the result of a search process that draws from existing knowledge and theories. The research should be rigorously evaluated and its contributions should be verifiable. Finally, the research should be communicated to the relevant audience in order to disseminate the resulting knowledge. Several models for conducting Design Science Research have been suggested which, taken together, formalize of a detailed research process as a methodology (Kasanen, Lukka and Siitonen, 1993; March, Smith and Smith, 1995; Lukka, 2003; Hevner *et al.*, 2004; Vaishnavi, Kuechler and Petter, 2005; Peffers, Tuunanen and Rothenberger, 2007; Hevner, A.; Chatterjee, 2010).

A process model should provide guidelines to assist researchers in understanding the requirements for effective DSR while at the same time recognizing the intellectual role

of scholar in making judgments on which guidelines should be applied to the research project (Hevner *et al.*, 2004). In this research, the Design Science Research Methodology (DSRM) of Peffers et al. (2007) was selected for the DSR process. The DSRM is an updated DSR process model, originally proposed by Hevner (2004), which includes six activities that function as guidelines for conducting DSR (graphically represented in Figure 10). The following paragraphs will summarize each of the activities of the DSRM by Peffers et al. (2007) and describe how they have been addressed in this research. Table 4 provides an overview of the research publication, their individual goals, and the methods applied.

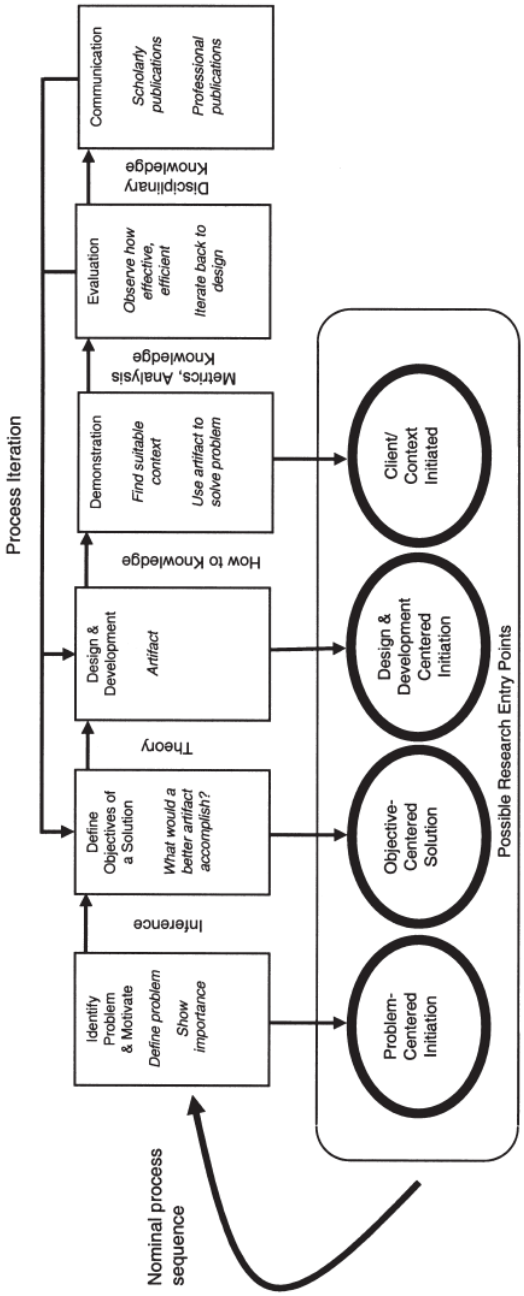


Figure 10. Design Science Research Method process model (Peffers et al., 2007, p. 10)

1. Problem identification and motivation: The specific research problem is defined in this activity and the added value of the solution is rationalized. Because the definition of the problem will be used to develop an artifact that can effectively provide a solution, it may be useful to atomize the problem conceptually so that the solution can capture its complexity.

In this study, Publications I and II cover this ground. The systematic review and bibliometric analysis of the published literature as well as the content coded involved in the publications reveal the need to consider SNSs when taking an ecosystem approach towards innovation studies. In this thesis, Section 1.1 and 1.2 also define research gaps, which are furthered as a major research objective and two consequent research questions in Section 1.3. In this study, the entry point to design science research is a problem-centered initiation (see Figure 10), and therefore the research process will proceed according to the identified problem.

2. Defining of objectives for a solution: This activity proceeds from the specification of the problem to the finding of its solution by considering those that are both possible and feasible. The objectives can either be quantitative and better than the existing situation, or qualitative or specify how the new artifact will support solutions to problems not hitherto addressed.

In this study, the requirements for viable solutions were derived from the research questions after visiting literature and evaluating system attributes of and their application in the ecosystem approach to innovation studies. Furthermore, a detailed discussion of using knowledge to connect a problem to its possible solution is in Chapter 2 of this thesis.

3. Design and development: This activity focuses on creating an artifact that are potentially constructs, models, methods, or instantiation or new properties of technical, social, and/or informational resources. Consequently, it is imperative to have knowledge of theory when moving from objectives to design and development when implementing a possible solution.

A novel method for adopting social network services data in an ecosystem approach to innovation studies was presented in Publications III, IV and V. Social network services data collection and analysis, the sentiment analysis of SNS data, crowd intelligence content evaluation and topic modeling are among the research methods used in these studies.

4. Demonstration: For this, the artifact is used to solve one or more instances of a problem. The artifact can be involved in an experiment, simulation, case study, or any other appropriate activity. It is also required that one demonstrates knowledge of how to use the artifact to solve the problem.

In this study, the demonstration of an artifact is included in Publications III, IV and V. The demonstration of the process was detailed in each publication, while Section 3.2 of the thesis summarizes the process.

5. Evaluation: This is to observe and evaluate how well the artifact actually solves the problem and involves comparing the objectives of a solution to the actual results recorded after the application of the artifact in the demonstration. Evaluation can take many forms, for example: comparing the functionality of the artifact with the solution objectives from activity 2, objective quantitative performance measures, or qualitative survey results from users of the artifact.

The required evaluation procedures were noted in Publication III where a machine learning model was designed to tackle a sentiment analysis task. The standard measures for calculating model relevancy and reliability, such as the F measure and receiver operating characteristic (ROC) curve are reported. Additionally, this thesis compares the evaluation with the objectives in section 5.

6. Communication: This activity of DSR requires that results are presented to the relevant audience and practicing professionals, when appropriate. Communication of the problem and its importance, the artifact, its utility and novelty, the rigor of experimental design, and its effectiveness are the main goals of this activity.

The results of these studies are communicated throughout the course of this study. Each of Publications I to V along with this thesis (when published), can be considered communicating results to the research community. Additionally, these results were communicated to the professional community at conferences, meetings at educational institutions, in social media, and in web publications.

Table 4 Research publications, goals, and methods

Research Publication	Research objective	Methods
Publication 1 *Emergence and Dissemination of Ecosystem Concept in Innovation Studies: A Systematic Literature Review Study*	To explore the core discussion of the "innovation ecosystem" and attempt to distinguish it from similar concepts	Systematic review and bibliometric analysis of the literature
Publication 2 *General System Theory Attributes in Innovation Ecosystem Research Landscape: A Bibliometric and Content Analysis of the Literature*	To break down the holistic ecosystem approach to innovation studies into recognized system attributes	Bibliometric and citation analysis of the literature. Performing content coding
Publication 3 *Brand Analysis in Social Network Services: Results from Content Analysis in Twitter Regarding the US Smartphone Market*	To operationalize SNS textual data to materialize the role of content and content producer in interactions observed	SNS data collection and analysis, sentiment analysis of SNS data, crowd content evaluation

Publication 4 *Efficiency Assessment of the Social Capital Capacity on Entrepreneurial Activity: A Perspective Driven From Social Media*	To develop metrics by incorporating SNS data to represent social capital in entrepreneurial-oriented activities	SNS data collection and analysis, sentiment analysis of SNS data
Publication 5 *Crowd Intelligence Participation in a Digital Ecosystem: A Systematic Process for Driving Insight from Social Network Services Data*	To witness the escalation of a digital ecosystem by observing how a mass intelligence crowd participates in an SNS during disastrous societal challenges	SNS data collection and analysis, topic modeling

The two major methodological approaches implemented during this research were: systematic review and bibliometric analysis of literature and social network services data collection and analytics. These two methods and their subsequent processes will be explained in this chapter. The final sub-section of this chapter will also describe the process in which a crowd evaluation survey was conducted for a content evaluation task.

3.1 Systematic review and bibliometric analysis of the literature

Due to the vast number of publications, there is a great need for methods that analyze, organize, and access information from large databases. As systematic reviews provide an overview of existing research and/or synthesize findings of meta data analyses (Krlev, Bund and Mildenberger, 2014), they can be used to identify, evaluate and integrate findings from high-quality studies that address one or more research questions (Baumeister and Leary., 1997; Cooper, 2003). Some important empirical approaches include Systematic Literature Review (SLR) (Kitchenham, 2004) and Systematic Mapping Study (SMS) (Petersen *et al.*, 2008). Systematic mapping studies in software engineering have been found to be effective methodologies as they adopt rigorous planning, follow repeatable and well-defined processes, and produce unbiased and evidence-based outcomes (Barn, Barat and Clark, 2017). The research in Publications I and II was conducted through a mixed methods systematic review which offers a comprehensive synthesis of two or more types of literature data (e.g. quantitative and qualitative) into a final, combined synthesis. The systematic mapping process, described in Petersen (2008), was broadly applied to these studies, but with some changes to suit the purposes of these studies. Guidelines for a systematic literature review from Kitchenham and Charters (2004) were used to search for relevant papers. Bibliometric data (quantitative measures used to assess research, in this case: publications and citations) analysis was utilized to collect the relevant literature which will be further described in a separate sub-topic. In Publication II, the full texts of shortlisted publications were considered for text segment extraction and coding after which the discussion of system attributes was manually observed.

3.1.1 Bibliometrics and citation analysis

Bibliometrics and citation network analyses were used to conduct a systematic literature review, with bibliometrics referring to the use of quantitative analyses and statistics to describe patters of publication within a given field or body of literature (Nicolaisen, 2010). Much has been written about bibliometrics and its use in measuring and exploring impact on a research field, set of researchers, or particular paper (Glänzel, 2015). For this research, bibliometrics was used to find literature relevant for performing an extensive qualitative content analysis. This bibliometric data collection and analysis was facilitated with the Network Analysis Interface for Literature Studies (NAILS)[3], designed in August 2015 by part of our research team (Knutas and Hajikhani *et al.*, 2015). NAILS is a cloud-based tool that performs statistical and Social Network Analysis (SNA) on bibliometric data (access at: http://nailsproject.net). The NAILS tools feature has demonstrated its ability to study the emergence of a research concept in scientific literature and the dissemination pattern of that concept in other disciplines (Hajikhani, 2017), patent portfolio comparative analysis (Ranaei *et al.*, 2016) and research topic analysis (Geissdoerfer *et al.*, 2017). Figure 11 illustrates the four steps of bibliometric analysis performed by NAILS (video instruction on how to use NAILS is accessible from: https://goo.gl/MuNgEG).

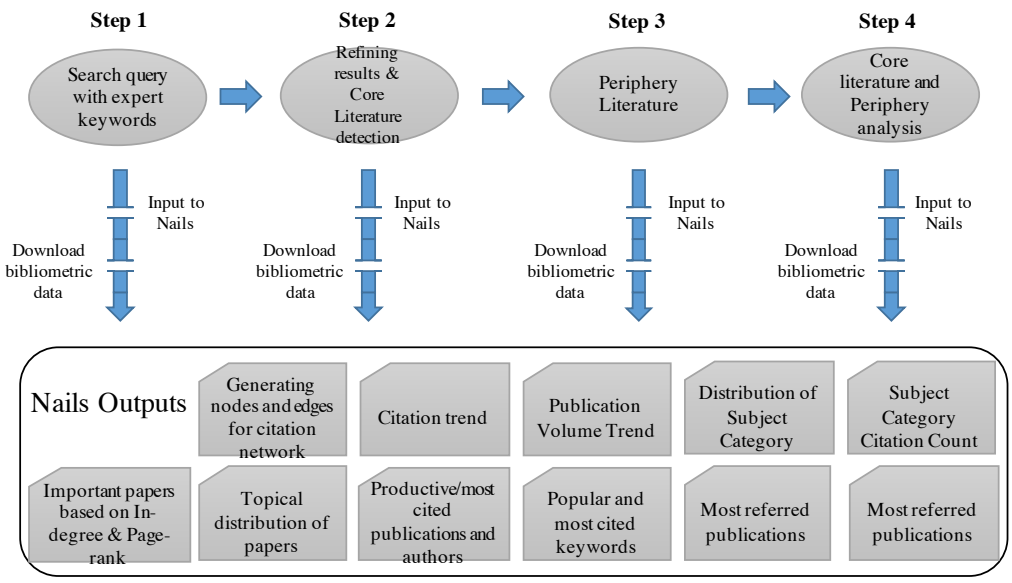


Figure 11. Steps to conduct bibliometric analysis with NAILS

A brief step-by-step guide to using NAILS is, as follows: **Step 1**, a set of keywords from the subject in question is collected to initiate the literature discovery phase. A search

[3] (http://hammer.nailsproject.net) online interface.

query with initial keywords and Boolean operators was constructed using the Web of Science (WoS) database for searching[4]. The bibliometric data is downloaded and bundled by a compression tool to be uploaded into NAILS; this task is repeated every time a new NAILS analysis is performed. Using the citation data, this tool generates a tailor-made report, providing abstract/keyword analyses, lists of productive authors/journals, and recommendations of top publications. **Step 2**, emphasis on refining the data collection process from the NAILS-generated report. The data refining process can be facilitated by the WOS interface, which makes the inclusion and exclusion of indexed publications both transparent and easy. The iterative emphasis in this process is on refining the initial data collection to identify the "core literature" from all of the literature in question. The bibliometric analysis and leveraging metrics, such as the in-degree (the number of citations of a paper sent to a directed graph) and PageRank (the number and quality of links to a paper counted to roughly estimate its importance) can help to identify the core literature. **Step 3**, delineates the perceptions in the core literature and uses these perceptions to help with identifying the most relevant literature in a multidisciplinary field of science. Analyzing the peripheral literature (papers citing the core literature) is helpful for exploring the dissemination of the research domain in question. **Step 4**, is concluding the iterative part of bibliometric data analysis by providing interpretable results that were generated by NAILS. NAILS reports can be interpreted in any of the stages of the process and some of the major outputs can be seen in Figure 11. The results can be interpreted or can assess publications in either the field under investigation or in another domain by looking at the number of citations and domain specifics. In order to explore the knowledge structure of a research domain, co-citation analysis of the references would become quite a helpful methodological approach (White and McCain, 1998). The previously mentioned technique also assists in the discovery of knowledge being diffused or influencing other research communities. This sheds light on networks of references, the social construction of a field, and its intellectual advances. Having data generated by NAILS show a network visualization of an author's citations is essential for comparing the network structure of collaboration among authors in any focused study.

With the methodological procedure of handling and analyzing bibliometrics data figured out, it was possible to give an overview on how the system and ecosystem concept was adopted in business and management studies (Publication I). A detailed comparison will be made to show these concepts within the context of innovation research. Later on, identifying the core literature for these two concepts will help us to understand the dissemination of these concepts into other domains.

3.1.2 Content coding and analysis

In Publication II, content analysis and bibliometric analysis are combined in an effort to identify the important literature and gaps in the previous research. Therefore, a full

[4] WoS is maintained by Thomson Reuters and has 151 million documents indexed (as of 9 July 2018). It is considered one of the most important databases for scientific bibliometric data among other competitors (i.e. Scopus, Sciencedirect and Google scholar).

content analysis was performed on the selected core literature for deeper review. A coding system was devised to signify the use of any of the system attributes defined in the previous section. The analysis was performed using the Maxqda software package[5]. Search queries were employed to identify which text segments to use. Maxqda analyzed the literature by search query and rules set, then generated text segments, identified co-occurrences of text, and provided insights into the occurrence patterns of coded segments. Figure 12, illustrates the content coding environment of the software which was utilized.

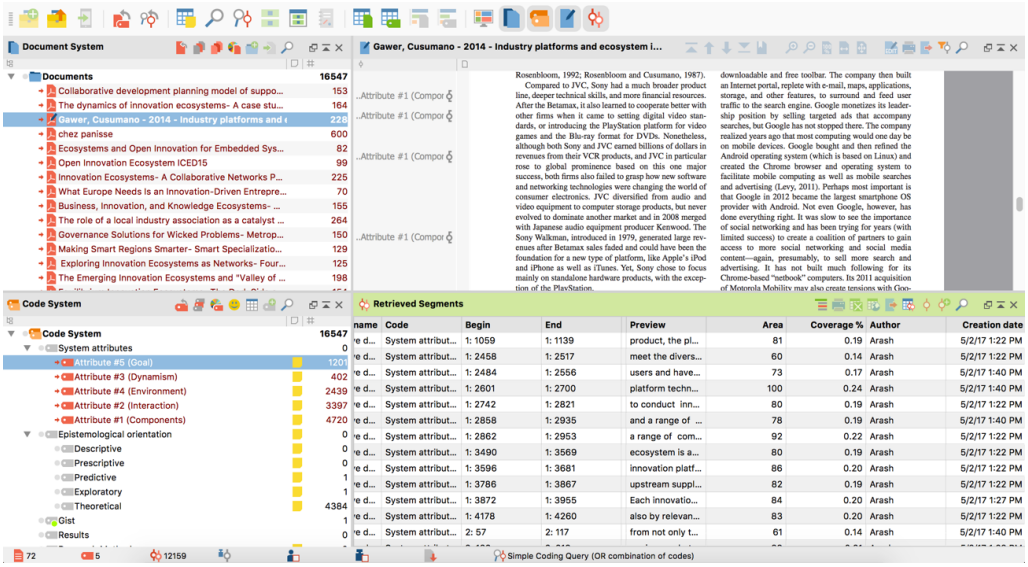


Figure 12. Content coding software interface

Referring to the Figure 12, on the top half of the software interface, there are selected publications (on left side) and the individual publication under content coding assignment (on the right side). The bottom part of the software interface are the tags associated to GST attributes (on the left side) and the extracted contend from the publication based on the tags in a tabular format (on the right side).

3.2 Social network services data collection and analysis

Internet data are available in various formats; social networks services provide one type/form of these data. In the early 20th century, prior to the availability of such data, sociologists interviewed people to understand their social connections and in doing so, formed small social networks for analysis (Anggarwal, 2011). Today, due to the activity on social networking platforms, such as Twitter, it is possible to do a study on SNSs in

[5] Maxqda 2018 [computer software]. Berlin, Germany: VERBI Software. Available from https://www.maxqda.com

real time thanks to the exceptionally large amount of content available and the millions of nodes and billions of edges.

In the past decade, the rise of computational power opened new opportunities for data analysis. At around the same time, an exponential growth in internet usage has increased the amount of data available. The ability to quickly access these multifaceted data along with the availability of ever-increasing computational power has led to rapid development in the field of social data analytics (Adedoyin-Olowe, Gaber and Stahl, 2014; Olshannikova *et al.*, 2017). Gartner (2017), a research and advisory firm on information technology, defines social data analytics as the analysis of people interacting in social contexts, often with data obtained from social networking services. The data in SNSs often comes unstructured, meaning it is neither organized nor presented in a pre-defined manner. That being said, while the data may contain dates, numbers and geo-locations, it is typically text-heavy. This study will take advantage of advancements in data mining and text analytics to search the SNSs for meaningful data. Taking a closer look at one such SNS, Twitter, can reveal its unique characteristics and features as a microblogging service, as is illustrated in Figure 13.

Post	**Body**	**Urls**
		Usermention
		Tweet Lang
		Media (Video, Picture)
		Hashtag
	Provider	**Followers Count**
		Followed by Count
		Profile Description
	Location	**Country/City /State**
	Like counts	
	Link	
	Retweet count	
	Posted time	

Figure 13. Twitter meta data illustration

In order to use Twitter as a data source, a multi-component semantic and linguistic framework was developed to collect, prepare, and analyze said data in the hopes of discovering meaningful information. Getting insights from SNSs was the main agenda of Publications III, IV and V, in particular. The overall architecture required to process data in SNSs is presented graphically in Figure 14. While Twitter (twitter.com) is one of many SNS data sources, this process is highly extent generalizable to most data in SNS platforms. The present process included three major phases—capture, curate and consume—each of which has two consequent sub-phases, as is shown in Figure 14.

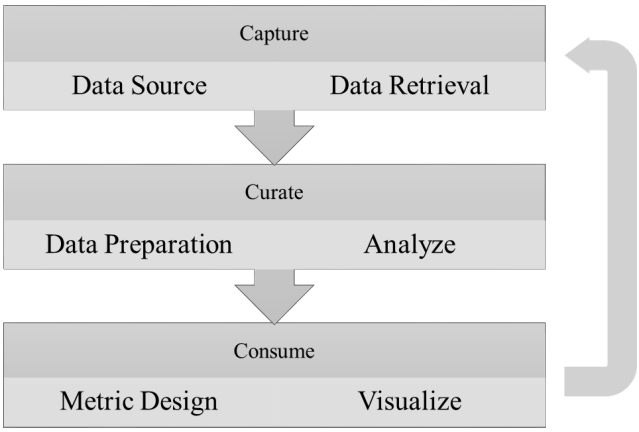


Figure 14. SNS systematic data analysis

Capture: This is the process of data collection which involves selecting the data source, searching for the data, and collecting it for other usage. The primary way to ensure that the content retrieved will be of any interest is by inputting a proper search query. Various specifications can be implemented, such as: keywords, length, date, etc., in order to target the topic of interest. In other words, the required data is obtained by a set of criteria embedded within the search query. Some SNS platforms, Twitter, for example, offer the possibility to retrieve data via live stream. Twitter provides application programming interfaces (APIs) to access tweets and information about users and posted content.

Curate: Data curation is a comprehensive term used to specify processes and activities related to the organization and integration of data collected from numerous sources. Data retrieval methods are often loosely controlled, resulting in out-of-range values. In order to avoid such values, data preparation is necessary for reducing the amount of irrelevant and redundant data in a collected set. This task is also imperative for normalizing the data for better knowledge discovery. Depending on the context of the study or expected results, data analysis can be very subjective, so two primary tasks for analysis—*data feature extraction* and *data classification*—are also needed. The intent to extract features to determine further distinctions and categorization of the data drives values (features) from the data related to the context of the knowledge discovery process. Classification of data occurs in order to reduce the dimensionality of the data. This is an approach derived from the general hypothesis of a knowledge discovery task to distinguish the data points from the mass that best fit the context. In order to understand the major discussions and the topic(s) they represent, topic modeling of both publication data and social network services data was performed for this study.

Consume: This refers to publishing information derived from data in a digestible format. Insights gained from the results can be presented in a visually appealing way or can be used as a metric that can be combined with other data points for further interpretation.

This research has benefitted from machine learning algorithms that can retrieve topical abstracts from publications, detect text polarity in social network service data, and approximate topical distance in SNS discussions over time. Originating in computer science and evolved from pattern recognition research, the automated process of categorization (or classification) of an object like text has created a growing interest in the utilization of machine learning (ML) practices. Due to the increased availability of documents in digital form, the need for flexible ways to access them has arisen (Sebastiani, 2001). Therefore, the activity of labelling natural language text (text classification or topic modeling) by applying machine learning algorithms to it has created an opportunity for processing large amount of text automatically and with greater insights. Next, the two ML approaches which were adopted for this thesis will be introduced in two separate subtopics.

3.2.1 Sentiment analysis

Finding polarity in public opinion from tweets was facilitated by an automated Sentiment Analysis (SA) technique. The purpose of this was to be able to automatically classify a tweet as positive or negative; this process had to be done quickly when working with such a big set of data. This implementation of the sentiment analysis of tweets using Python [6] and the Natural Language Toolkit (NLTK)[7] was adopted from Luce (2012). To understand the perceived sentiment of a tweet, a corpus of tweet ratings was needed. The computer, however, needed a training corpus and/or documents with information to learn from. The more documents the computer had, the more accurate the results of its ratings. In sentiment analysis, the training corpus always has example documents that are manually annotated into categories. Having learned from example, the computer can apply the acquired knowledge to new documents (a hold-out corpus or training set) and classify them then into sentiment categories. In order to construct a training set for this study, a multi domain dataset of Blitzer et al. (2007) was utilized. The database composed of product reviews from six Amazon product domains (e.g., book, DVD, electronics, kitchen, music, and video) which were manually annotated and classified as either positive or negative. Instructions from Luce (2012) were followed in order to create a classifier by extracting and labeling the relevant word features. The constructed variable "training set" includes the labeled feature sets which contain a list of tuples with each tuple containing the feature dictionary and the sentiment string embedded in it. Now, the training set can train the classifier to predict the opinion polarity of unseen tweets. Figure 15 is an illustration of the steps for conducting the sentiment analysis.

[6] Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language available at http://www.python.org

[7] http://www.nltk.org/

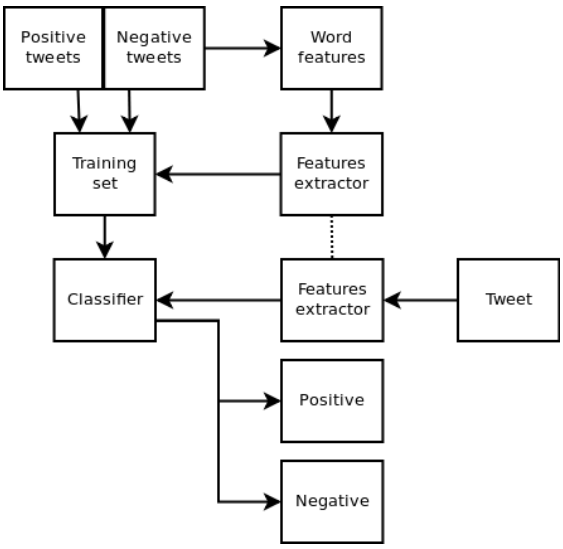


Figure 15. The sentiment analysis steps (adapted from Luce 2012)

Although there are many classification methods for performing sentiment analysis (e.g. Naive Bayes, Maximum Entropy, SVM), Naive Bayes is the algorithm that was used in this research as it enables the classification of documents into positive and negative categories. Naive Bayes is also a probabilistic model which works well on text categorization (Weikum, 2002).

3.2.2 Topic modeling

In machine learning and natural language processing, topic modeling is a technique used to discover the abstract "topics" that occur in a set of documents. Topic modeling is one of the most powerful techniques for analyzing large volumes of unlabeled text. Topic modeling approaches have been used in areas such as intellectual property research (Ranaei, Suominen and Dedehayir, 2017), political science, (Grimmer, 2009) and bibliometrics (Gerrish, 2010).

In Publication I, a topic modeling technique was adopted to analyze the contents of the publications' abstracts. This technique was useful for discovering the abstract "topics" that appear in a large collection of documents in order to explore hidden semantic structures in a body of text. The "Latent Dirichlet Allocation (LDA)" introduced by Sievert and colleagues (2014) was used to perform topic generation on the abstracts analyzed. A visual representation of both the innovation ecosystem and innovation systems in Publication I showed the major topics under discussion for each concept. The topic modeling resulted in a visual form which aided in understanding of both concepts, as one could see where the concepts overlapped as well as the number and distance of concepts relative to each other. In Publication V, a topic modeling technique was utilized

to explore the formation of a discussion on Twitter over time regarding the Fukushima incident. In this experiment, LDA and the consequent visualization toolkit developed for that purpose visually showed the periodic Twitter discussion of topics over time (Blei and Blei, 2003). This approach enabled the measurement of intertopic distances of identified topics (represented as circles) onto a two-dimensional map and the change those distances over time.

3.2.3 Crowd intelligence content evaluation

The task of evaluating the likelihood of a particular piece of content being read was handled via survey. In order to investigate which type of tweet gets rated higher by users, a content evaluation survey was designed for Publication III. After first creating a list of top producers consisting of the profiles with the highest number of followers, focus was directed to the top tweets from these top producers. With the help of internet-based tools, the survey creation process was outsourced. Technically speaking, the concept of c*rowdsourcing,* defined as the practice of getting ideas and services from a large group of people willing to contribute their input, was applied for the purposes of this study. This type of outsourcing can be applied to various activities, most notably to tedious tasks (Safire, 2009).

In an effort to measure the level of the user interest for these top tweets, a survey asking the likelihood of their reading the content in question was designed. Mechanical Turk (MTurk), a service from Amazon, was used to perform this content evaluation task. A crowdsourcing internet marketplace, MTurk allows individuals and businesses (known as requesters) to coordinate the input of individuals to perform tasks that cannot be automatized by computers. Tweets are notoriously hard to classify given all the abbreviations, sarcasm, and hash tags; therefore, human subjects would provide the best results. Participants were asked to evaluate their interest in reading the content on a Likert scale (1=not at all interested to 5=very interested). 300 people answered the questionnaire and evaluated tweets for content quality. Each participant was given a randomly selected question (out of 120 total questions) for the evaluation task. Figure 16 shows a sample question.

Please specify the level of your interest for reading the 5 Tweets option provided.

	Not at all interested	Not very interested	Neutral	Somewhat interested	Very interested
Option 1	○	○	○	○	○
Option 2	○	○	○	○	○
Option 3	○	○	○	○	○
Option 4	○	○	○	○	○
Option 5	○	○	○	○	○

Figure 16. Crowd content evaluation survey

The crowd evaluation of tweet's content readability interest had been transformed into a variable which then was used to explore the variability in content producer type and users readability interest.

4 SUMMARY OF PUBLICATIONS

This section of the dissertation summarizes the objectives and contributions of the five publications on which this thesis is built. Although these discussions remain high-level, for more detailed discussions, one should refer to the original publications.

This section reviews the objectives, results, and main contributions of each publication included in this study. Each publication made contributions that, taken together, answer the main research questions and also the objective of this dissertation.

4.1 Publication I:

Emergence and Dissemination of Ecosystem Concept in Innovation Studies: A Systematic Literature Review Study

4.1.1 Background and objective

The objective of this publication was to investigate the abundant amount of academic literature on the ecosystem approach which had attracted the attention of many researchers in several fields, such as: business and management, innovation, and entrepreneurship. The term "ecosystem" originated from ecology, which considers living organisms and their interaction with the environment they inhabit (Papaioannou, Wield and Chataway, 2009). The application of this biological concept to economic studies was endorsed by Rothschild (1990) and later applied to business and management (B&M) studies by Moore (1993). Biology inspired the adoption of the ecosystem approach in B&M-oriented studies as the interaction between organizations, their environment, competition, and forces that maintain stability while spawning changes seem to parallel biological concepts (Rothschild, 1990; Moore, 1993). This multiplicity of research directions created excitement for the adoption of this terminology in academic literature, especially in B&M studies. On account of this scholar's interest in adopting this novel and emerging concept, the challenge of settling on accepted definitions from past discussion was presented (Oh *et al.*, 2016). The validity of the ecosystem analogy in innovation systems was critiqued; for example, Oh et al. (2016) suggested that the concept of the innovation ecosystem began to infiltrate spaces more traditionally described as innovation systems, triple-helixes, or clusters which led to ambiguity in the usage and application of this concept. A debate on the validity and value of the ecosystem approach to innovation is still ongoing (Oh *et al.*, 2016; Ritala and Almpanopoulou, 2017). There appears to be no consensus on this topic in the academic literature. The take-away from this debate is that too many unconnected concepts made understanding the concept difficult if not impossible.

This publication tackles the research question of how the "innovation ecosystem" differs in comparison to its predecessor: systems innovation. This paper employs publication bibliometric data to perform citation analysis to observe knowledge creation and the diffusion of the emerging concept of ecosystem in innovation studies. Hitherto,

bibliometric data analysis does not deliver direct insight on future developments (Small, 1988; McCain, 1990), but results will explore the evolutionary path of the ecosystem concept in B&M studies.

4.1.2 Results and main contributions

The bibliometric data analysis was enabled by the help of a cloud-based tool for Network Analysis Interface for Literature Studies—"NAILS"—that was developed and released (Knutas and Hajikhani *et al.*, 2015) alongside this research. Unlike the systems concept, the adoption of the "ecosystem" concept occurred without consistency or interconnectivity of authors. One possible explanation for the lack of author collaboration might be because some believed using this new terminology would eventually create a disconnect with older and/or overlapping earlier concepts. As a result, the paper emphasized the importance of developing an ecosystem vocabulary that is commonly understood and consents to a comparison among studies. More case studies to illustrate the usage of concepts are necessary to clearly differentiate between them. A comparison of case studies looking into innovation system vs innovation ecosystem would help to clearly differentiate these two concepts. The ecosystem approach to innovation studies certainly needed to identify theoretical approaches, principles, indices, models, frameworks, and tools to create a mutual point of reference. The development of these approaches would form the necessary foundation for future empirical research as well as theory development and validation (Oh *et al.*, 2016). In this study, it was concluded that the ecosystem approach to innovation studies matured through a number of publications whereas the attachment to an epistemological orientation had not been solidified. It is hoped that this review invites researchers to initiate more rigorous research that helps to expand the understanding of the ecosystem approach to innovation studies.

4.2 Publication II:

General System Theory Attributes in Innovation Ecosystem Research Landscape: A Bibliometric and Content Analysis of the Literature

4.2.1 Background and objective

The innovation ecosystem was introduced to business and management (B&M) perspectives as a network of collaborative components competing with and complementing each other for a value proposition (e.g. (Adner, 2006; Adner and Kapoor, 2010; Ritala *et al.*, 2013)). One of the major discussion points highlighted by the ecosystem approach to innovation is the complex systems or network effect, an inspiration which derives from the biological origins of the ecosystem term (Russell and Smorodinskaya, 2018). The systems approach, in general, assists with studying the functions of complex organizations and is utilized as the basis for new kinds of approaches such as "ecosystem," for more holistic descriptive power (Anderson, 1999; Peltoniemi and Vuori, 2004; Mele, Pels and Polese, 2010). Foundational to systems thinking was Bertalanffy's introductory work on the General Systems Theory (GST)

(Ludwig Von Bertalanffy, 1972). Systems theory is a theoretical perspective that analyzes a phenomenon as a whole and not as simply the sum of elementary parts, focusing on the interactions and on the relationships between parts in order to understand an entity's organization, function, and outcomes (Mele, Pels and Polese, 2010). In order to reduce the conceptual ambiguity of the innovation ecosystem concept, this study uses GST as a lens to evaluate the literature. In general, systems thinking is the cognitive process of studying and understanding systems of every kind. This framework is increasingly being used to tackle problems in a wide variety of disciplines, such as: business, management and innovation studies (Jenkins and Youle, 1968; Mele, Pels and Polese, 2010; Johannessen, 2013). In order to make an in-depth analysis of the ecosystem approach to innovation studies, the main research question of this publication was formulated, as follows: "How has the systems approach and its attributes been adopted in the innovation ecosystem literature?" Based on a review of the GST literature, five major attributes for the system were categorized as: Attribute #1 (Components), Attribute #2 (Interaction), Attribute #3 (Dynamism), Attribute #4 (Environment), Attribute #5 (Goal). By introducing the system's main attributes, it is possible to examine and detect these attributes in the publications that were part of this study. The availability of scientific papers as full text and in machine-readable formats provided an opportunity for a more in-depth textual analysis of full papers.

The attributes that were extracted can be seen in the seminal works on innovation ecosystems by Adner (2006 and 2010). Firms and end customers have explicitly used the term *Components* to refer to either complementors or intermediaries. *Interaction* is a part of a collaborative network and/or collaborative arrangements in the context of reducing the cost of coordination for a firm. Whereas, *Dynamism* has been described as an evolutionary process within the ecosystem where a process takes place iteratively as an assessment for the target, that being maximizing the value proposition of the ecosystem. *Environment* is accounted for by focused case studies. Finally, *Goal* is reflected in the context of studying risk assessment and expectation alignment. This study suggests that systems theory and its attributes can provide a powerful perspective and methodological lens for the analysis of the newly-emerging trend of using an ecosystem approach to innovation studies. Therefore, by introducing a system's main attributes, it is possible to examine and detect the attributes in the focus literature.

4.2.2 Results and main contributions

Bibliometric data analysis was done with the help of the NAILS tool which targeted pieces of literature which were relevant for further qualitative content analysis. The automated text segmentation analysis processed the full text of a publication in order to detect the five systems attributes. The results showed that more attention to be paid to the *Dynamism* and *Goal* attributes. Dynamism was a holistic glance at solutions which result from collaborative innovation that recognizes the role of inputs. The notion of Goal as a system attribute considers the structure of the whole system as a set in order to produce a certain goal as well as its justification. This is often called the purpose, outcome, or goal of a system. Dynamism and Goal appear in the current literature in case studies with

particular dynamic behavior and shared goals. However, on a theoretical level, a shared goal and a somewhat implicit dynamism of behavior among components of a system should be expected. Within business and strategy studies, we can expect that actors make strategic decisions to act in an ecosystem. The actors' capability to manage their position and that of others in the ecosystem governs their ability to create and capture value from participation (Hannah, 2013). Current literature, however, suggests that actors operate towards their own goals and that there is now a shared understanding of dynamism or goals with an ecosystem (Jacobides, Veloso and Wolter, 2014). This assumption requires more attention, as being able to change and share a goal seem central to the ecosystem's health—a key indicator of "well-being, longevity and performance" of the ecosystem (Hyrynsalmi *et al.*, 2012). Also, an attempt was made to bridge the complexity of a system by network visualization (Russell *et al.*, 2011). Network visualization and social network metrics adoption have benefited the community by documenting the evolution of the ecosystem concept. Network visualization can provide evidence about ecosystem transformation and opportunities for orchestrating this transformation (Still *et al.*, 2014). Social Network Analysis has been utilized for better understanding of this ecosystem transformation. Visual representation of social networks is important to understanding the network data and conveying the result of the analysis (Huhtamäki and Rubens, 2016). From a theoretical perspective, this paper contributes to the overall understanding of the adoption of the ecosystem approach in innovation literature and the areas where the approach should be improved or further studied. From a methodological perspective, this study provides an innovative method for systematic in-depth literature review work. It demonstrated the applicability of a state-of-the-art quantitative analysis as a complement to traditional qualitative methods of reviewing the literature which is quite helpful for tasks such as concept evolution and knowledge discovery.

To conclude, the results provide interesting implications to further emphasize systems component usage in innovation ecosystem literature. General systems theory has been advanced from a conceptual framework to one with a more solid toolkit. This work invites other scholars to contribute to the evolution and development of the systemic nature of the ecosystem concept within innovation studies. Models and tools that can simplify the complexity of social and economic exchange in a meaningful way without eliminating the richness that a solid traditional system approach provides are needed. The resources for these models and tools can be found through a collaborative effort from diverse academic disciplines, which can provide the cross-fertilization that is needed for the next step of analysis.

4.3 Publication III:

Brand Analysis in Social Network Services: Results from Content Analysis in Twitter Regarding the US Smartphone Market

4.3.1 Background and objective

The literature review presented earlier in this thesis (Chapter 2) emphasized that innovation can be reduced to tangible and intangible, or soft, characteristics, with intangibility affiliated with business models, networking, and brand innovation. Following the investigation of the ecosystem approach to innovation and the characteristics attributed to it (Publications I and II), this study intended to materialize these intangible attributes. This study operationalized SNS data as a major component of an innovation ecosystem to give insights concerning public opinion regarding a brand presence. Despite being a time consuming, labor intensive and expensive method for gathering data, SNSs offer fast and relatively cheap access to mass amounts of data which is mostly publicly available and accessible to researchers. The wisdom of the crowd, documented on the SNS, apparently plays a key role in major decision making in almost any context (Nair, 2011). Therefore, finding a practical way to explore and mine valuable information from user-generated content (UGC) data is essential. In this study, Twitter data was utilized due to the platform's popularity and the fact that its structure was suitable for conducting this experiment. Due to the platform of the chosen data source, a wide variety of research issues in mining Twitter data had to be investigated. Several studies have explored extracting public sentiment (Pang, Lee and Vaithyanathan, 2002; Das and Chen, 2007; Pang and Lee, 2008; Go, Bhayani and Huang, 2009) in a variety of cases. However, by adding another important dimension, the content generator's impact on grabbing the public (users') attention, this study took a deeper look than previous studies. This dimension was added based on the fact that certain criteria plays an important role for whether a tweet is being read and distributed. Thus, the main research question presented was: "How does content quality perceived by the user of an SNS vary among different types of content generators?" To demonstrate computer advancement in textual analysis of SNS data, a case study was utilized to capture discussions related to smartphone manufacturers in the US for the duration of one year. The data was retrieved from Twitter and a machine learning model was design and implemented to assign sentimentality (positive or negative) to tweets. This effort was done as part of the research sub-question: "How does public sentiment reflect content sentiment polarity within different categories of content providers in SNSs?" Additionally, in order to capture the likelihood that Twitter content was being read, a survey was designed to evaluate crowd intelligence regarding the content of tweets. Finally, multiple variables were constructed and a correlation analysis was executed in order to explore possible relationships. As a result, a number of propositions for a firm's social media marketing strategies shall be developed.

4.3.2 Results and main contributions

This research extended our understanding of content impact in social network services (SNSs) by collecting SNS data and creating indices and variables to test the hypothesis in various dimensions of content and content producer and on the sentimentality of content consumers. In order to investigate techniques for generating indices to create meaningful assumptions for testing, a case study was acquired for this experiment. This

in-depth analysis of SNS data revealed the distribution of detected content generated in SNS profiles is different for each company.

The results show that the content in certain profiles (professional user and purely personal) were valued higher and read more often than the other three profiles (corporate and business, feed/news, and viral/marketing). Additionally, the case study done for this publication revealed the relationship between the types of content producers in SNS, relatively, the type of content they produce, to the level of interaction they receive. In this publication, the aggregate view on SNS content sentimentality was challenged through a study of the content distribution in five different profile categories. The Pearson's correlation analysis was applied to study the relationship between variables. The results impose the different carrying capacity of negative and positive highly-impactful content, in this case: tweets. The results show that the weight and value of messages with negative sentiment were higher compared with the positive and neutral ones. This negative bias can be somewhat explained by the fact that the content was generated by the personal and professional user categories.

Overall, this study demonstrated the importance of the content representing companies in SNSs and emphasized the importance of content generator and the negativity affiliated with the interaction with content. The advanced textual analysis method utilized on SNS data (sentiment analysis) revealed that that public sentiment was projected onto a company's product. At the same time, crowd intelligence was ascertained by evaluating the top tweets of each company, as the importance of the content producer and his or her intent to share it on SNSs was also recognized. In particular, these findings comply with the negativity bias also known as the "negativity effect theory." This idea states that, even when of equal intensity, things of a more negative nature (e.g. unpleasant thoughts, emotions or social interaction; traumatic events) have a greater effect on a person's psychological state and processes than neutral or positive ones (Lewicka, Czapinski and Peeters, 1992; Baumeister *et al.*, 2001; Rozin and Royzman, 2001).

The value of this study in large extent lies in its attempt to materialize SNS data for further in-depth investigation on this seldom explored data type. The alignment of this study with this thesis' agenda is that the importance of operationalizing SNS data was discussed when advancements in computer science and text analysis techniques were highlighted.

4.4 Publication IV:

Efficiency Assessment of the Social Capital Capacity on Entrepreneurial Activity: A Perspective Driven from Social Media

4.4.1 Background and objective

The holistic and systemic perspective of an economy's performance in innovation and entrepreneurship raises the question of, which is the most influential of the many measures introduced. The previous simplified linear models of explaining innovation by

using indicators such as research and development (R&D) expenditure or patenting is no longer adequate (European Commission, 2013). In innovation evaluation and measurement practices, soft aspects of innovation have received more attention in the past decade (Grimaldi *et al.*, 2017). Intangible, or soft, aspects of innovation have been discussed using different terms, such as human capital, social capital, and culture (Ralph H. Kilmann, Mary J. Saxton, 1987; Meyer and Harper, 2005). In the "quadruple helix" literature proposed by Carayannis and Campbell (2009), focus is placed on social capacity. In the fourth helix, particular attention is put on highlighting the importance of human capital and, at large, social capital in fostering innovation. While it is important to encounter the social and human capital capacity into an analysis, it is difficult to capture the influence of society on innovation and entrepreneurial activity. The literature has discussed the importance of computer-supported communication, such as social network services and the concept of social capital (Kobayashi, Ikeda and Miyata, 2006; Carmichael, Archibald and Lund, 2015). Therefore, the main research question was set forth as: "How can social network services data be operationalized to evaluate the social capital capacity of an economy with regards to entrepreneurial-oriented activities?" This research question was tackled by adopting a systemic procedure for capturing and identifying entrepreneurial-oriented discussions on social media within the geographical boundaries of the countries under study. The follow-up research question of this study set out to evaluate the efficiency of an economy based on its social capacity for entrepreneurial-oriented activities.

This study adopts both quantitative and qualitative methods, tackling the complex issue of countries' entrepreneurial-oriented activity through indicators both new and traditional, which enabled a benchmarking through analysis of results. Special focus was placed on social network services in order to capture entrepreneurial activity as well as receive established reports on the country-level performance benchmarking practices used to measure innovation and entrepreneurship. Twitter was selected as the SNSs for this study, and because of the context of the study, the startup ecosystem of each country was targeted in order to capture entrepreneurial-related discussion. The ratio indicating a country's entrepreneurial activity on Twitter was normalized according to population size and internet penetration rate for each country. This ratio illustrates that a country achieved its capacity for entrepreneurial-oriented discussion on social media. The results of this experiment serve as a proxy for the entrepreneurial activity of the countries studied. These numbers will be used to calculate the efficiency ratio of countries utilizing their social capital resources towards entrepreneurial-oriented activities. The level of analysis was performed at the country-level by looking at a sample of European countries. The essence of performance analysis inherent in benchmarking was leveraged for this study. This idea of a comparison between units of analysis by relative efficiency scores is found in the performance analysis literature (Selden and Sowa, 2003; Greiling, 2006; Scott and Davis, 2007). In order to assess the efficiency of input efforts transforming into output, a non-parametric method dominant in operation research and economics, known as Data Envelopment Analysis (DEA), was utilized. Furthermore, process of converting variables into a model was run via DEA for the purposes of this study.

4.4.2 Results and main contributions

This paper attempted to capture the capacity for social capital and illustrate its indirect effect on entrepreneurship-oriented activities. The study also discussed the importance of social network services outlets as the pulse of society and their dominance in hosting a large portion of societal discussion. Furthermore, SNS data was leveraged in order to capture the intensity of the discussion about startup and entrepreneurial activity on the SNS platform. The countries were ranked simplifying the relationship between set of items by assigning a sequence of ordinal numbers. This study proceeded to model an input-output process which generated the efficiency ratio that benchmarked the countries' standing where societal capacity for entrepreneurship was concerned. Therefore, utilizing the DEA model was necessary for constructing and calculating the countries' efficiency ratio. The simplified sketch of input and output variables is presented in Figure 17.

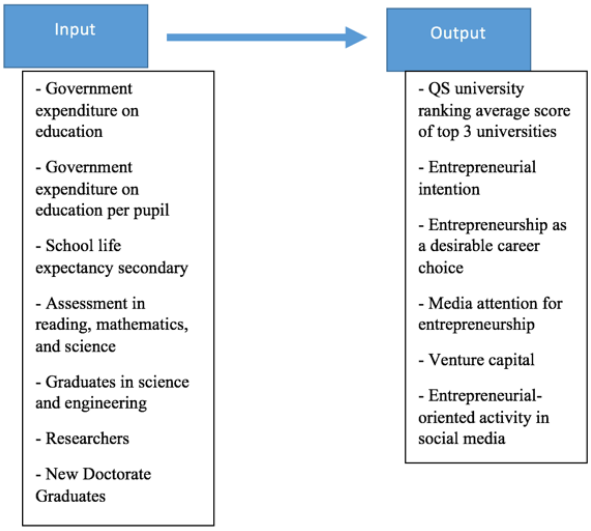


Figure 17. DEA input – output model

Efficiency measures were calculated which generated a new scale and ranking accordingly, with an emphasis on the efficiency of social capital capacity rather than proficiency. The efficiency score was calculated in such a way that countries that share a similar mix of inputs would be compared which each other; this translates to similar possibilities for efficiency. The main objectives of DEA is to measure the efficiency of a Decision Making Unit (DMU)s on a scalar measure, ranging between zero (the worst) and one (the best). The efficiency state of DMU in DEA analysis is described as when further improvement will not be achieved without harming some other input or output. Therefore, in the context of this study, efficient countries were the ones that utilized their capacity more productively than less efficient ones. The efficiency scores are estimated using a variable return to scale, output-oriented DEA with 7 inputs and 6 outputs. Decision making units (DMUs) with efficiency scores equal to 1 are efficient, while

DMUs with scores greater or less than 1 are inefficient. Two models have been calculated for their efficiency scores: CCR based on Charnes et al. (1978) and BCC proposed by Banker et al. (1984). Meanwhile, it is important to note that the efficiency scores are relative scores. A high relative efficiency score does not mean that there is no room for improving performance. The results of these evaluations are displayed below in Table 5.

Table 5 DEA analysis BCC and CCR efficiency scores

BCC Output DMU	*Score*	*CCR Output DMU*	*Score*
Ireland	1	Ireland	1
Netherlands	1	Netherlands	1
Luxembourg	1	Luxembourg	1
Hungary	1	Hungary	1
Portugal	1	Portugal	1
Poland	1	Poland	1
Slovakia	1	Slovakia	1
Bulgaria	1	Bulgaria	1
Latvia	1	Latvia	1
Romania	1	Romania	1
United Kingdom	1	Croatia	0.998495
Finland	1	Greece	0.99245
Croatia	1	Italy	0.96797
Greece	1	Finland	0.959859
Slovenia	0.995086	United Kingdom	0.954028
Italy	0.969573	Belgium	0.943379
Germany	0.955024	Slovenia	0.92143
Belgium	0.946866	Germany	0.905398
Sweden	0.90235	Sweden	0.879499
Estonia	0.868719	Estonia	0.829102
Spain	0.723702	Spain	0.706288

Overall, this publication employed a meso-level approach, where country was used for the purposes of leveraging SNS data in order to capture social-economical aspects of society. This study and its constructed model has the unique ability to examine the efficiency of countries in utilizing their social capital resources towards entrepreneurial-oriented activities. The efficient countries, according to the model, had a better balance of social capital and entrepreneurial-oriented activity on SNSs when compared to the less efficient ones.

4.5 **Publication V:**

Crowd Intelligence Participation in a Digital Ecosystem: A Systematic Process for Driving Insight from Social Network Services Data

4.5.1 Background and objective

This study will look at social network services (SNSs) within the transformation process of societies moving towards digitalization. The convergence of extremely large data sets,

known as "big data," generated by technologies, such as social media, has led to an unprecedented amount of digitalization currently fueling innovation in business and society (Legner *et al.*, 2017). The fast progress of technology has resulted in societies and communities that are connected and communicating on platforms such as social media or social network services (SNSs). The influence of SNSs on political and social issues is only getting greater and greater (Eom *et al.*, 2015). The corner stone of discussion within the study of SNSs is the possibility to facilitate the exchange of knowledge by sharing information quickly, globally, and to large numbers of individuals (Powell, Koput and Smith-Doerr, 1996). A recent assessment made by the World Economic Forum (2017), recognized SNSs as one of the forces driving transformational change in economies, industries, and global issues.

SNSs has been a dominant venue where people either participate in or passively consume events as they unfold. Recently, Twitter has been used to spread news and updates around the world and has been shown to have applications in natural disaster emergency situations, such as: earthquakes, floods, hurricanes, and wildfires (Hughes and Palen, 2009; Kireyev, Palen and Anderson, 2009; Starbird *et al.*, 2010; Vieweg *et al.*, 2010; Muralidharan *et al.*, 2011). SNS platforms allow for multidirectional network communication which can aid officials during disasters in compiling lists of the injured and the deceased. Not only that, but they can also contact family and friends of victims—all while connecting and organizing both casualties and responders (Cooper *et al.*, 2015). Furthermore, Twitter was shown to have the potential to increase survival during tornado-related disasters (Lindsay, 2011). SNSs have been increasingly used for building and supporting communities and affording self-expression and identity construction for individuals in the communities they belong to (Jaeger *et al.*, 2007). The capability of SNSs has been leveraged to initiate real-time information networks powered by communities and authorities.

The social infrastructure, such as intellectual and social capital, presented in SNSs is indispensable to digital platforms as it allows for "connecting people and creating relationships" (Albino, Berardi and Dangelico, 2015). Information and Communications Technology (ICT) also offer new avenues for openness by providing access to SNS content created through the interaction of users via highly accessible Web-based technologies. SNSs can refer to both the enabling tools and technology as well as the content generated by them. As SNSs are integral to the applications mentioned, this research argues that SNSs play a major role in current digital platforms and tend towards an explorative view over the discussions they host during disasters. Accordingly, the main research question became: "How can the role of social network services (SNSs) as a major component in digital platforms be materialized?".

This study proposed a systematic way to gain insights from SNSs data to better understand unification and cohesion in discussions on SNSs. Furthermore, for a better handle on the systematic approach to SNS data, an experiment regarding the Fukushima incident was conducted. The systematic SNS data analysis was replicated to collect data from Twitter to observe the topical evolution of the discussion over time.

4.5.2 Results and main contributions

In order to retrieve relevant data on the Fukushima incident, a search query was constructed to initiate Application Programming Interface (API) in Twitter. A systematic process was designed and utilized to get the required information from the collected SNSs data. The process is composed and presented graphically in Figure 18.

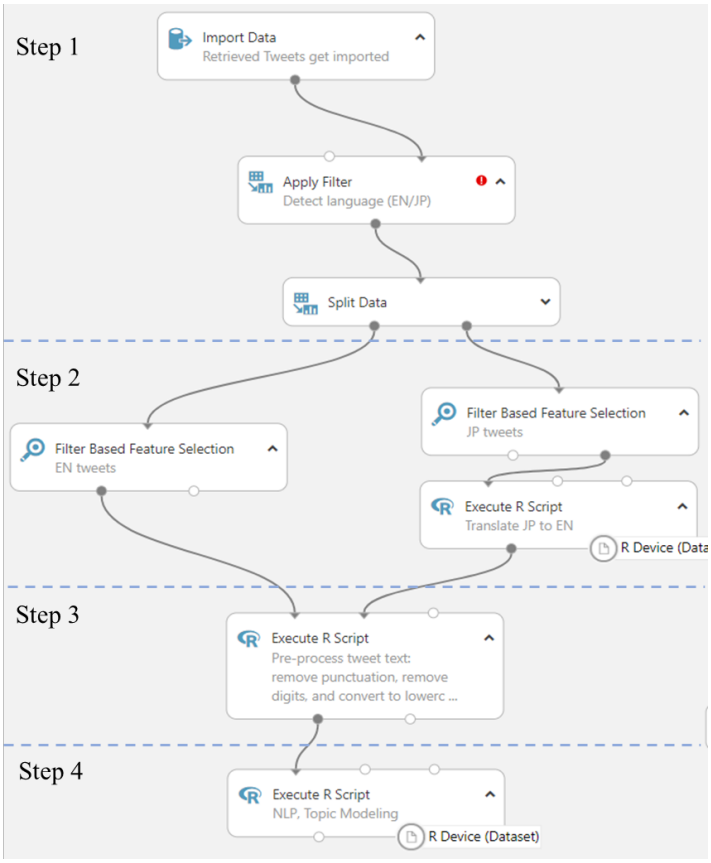


Figure 18. Twitter content analysis with Azure Cloud Computing Platform

In general, this study intended to learn about hidden similarities in discussion topics. Topic modeling uses a Nonparametric Bayesian Model to measure the similarity in documents and generates the topics. There are many techniques that are used to obtain topic models, in this study the Latent Dirichlet Allocation (LDA) was utilized. LDA was used as a probabilistic model for classifying tweets. The validation accuracy was maximized when there were 15 LDA topics. In other words, this model is best used for explaining the distribution of discussion topics when clustering is set for 15 topics. Topic models are generative, which means that they model texts as if they were generated from a certain probability distribution. The motivation behind this analysis was to observe the

distribution of topics of discussion in an SNS and examine the topical proximity as time evolves. Regarding the analysis of tweets from the Fukushima incident, after data retrieval, processing, and topical analysis, the matrixes presented in Figure 19 is one application showing the probabilistic distribution of discussion topics generated, also explaining the most variance in the data.

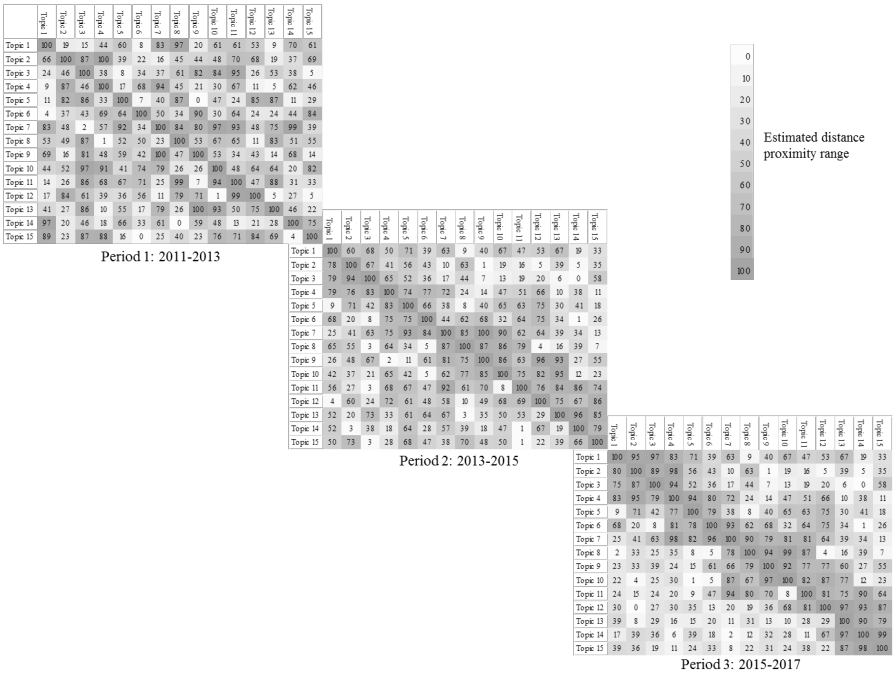


Figure 19. Estimated average distance between discussions discrete topic over three time periods

The trend was plotted for three separate time periods (2011-2013, 2013-2015, and 2015-2017) to enable comparison of topical evolution on SNS discussions. This analysis shows that the topical structure of the discussion progressed in close proximity to a third period from a more topical distance at the beginning of the discussion in Period 1. It is clearly visible that the discussions' topical coordination merged closer to a kind of diagonal position in the matrix in the third period; while in the first period, the fifteen detected topical discussions were positioned with a shorter estimated distance from one another. The periodic analysis of topical distance in major discussions on SNS platforms identified a pattern of collective knowledge formation and dissemination.

The contribution of this study to the field is yet another exemplification of exploiting SNSs data. This systematic methodological process enabled the efficiently and effectively processing of SNS data for insightful information which opens the door to including an SNS component to digital platform studies. Understanding these newly-emerging SNSs can help to define the footprint or scaffolding for making a progressive, dynamic framework that will steer towards the achievement of digital objectives. This agenda

further defines the advancements in computational capabilities necessary for developing accurate competencies of SNS textual content.

4.6 Summary

Overall, the five publications contributed to an understanding of social network services, their positioning in an ecosystem approach to innovation, and the way to operationalize their data. For a summary of the papers' contributions, see Table 6.

Table 6 Summary of the five publications in this thesis

Publication	Research objective	Main contribution
I: *Emergence and Dissemination of Ecosystem Concept in Innovation Studies: A Systematic Literature Review Study*	To understand the core discussion around the ecosystem concept in innovation studies	Furthering the academic discussion that promotes the use of the ecosystem concept in innovation studies
II: *General System Theory Attributes in Innovation Ecosystem Research Landscape: A Bibliometric and Content Analysis of the Literature*	To study the system attributes in an ecosystem approach to innovation studies	Exploring the presence of system attributes in the core innovation ecosystem literature. Adding to the literature by arguing that a lack of dynamism and stating of the goal of a discussion in the extended literature reduced coherency and furthering the extension of the domain to other disciplines.
III: *Advanced Methods: Operationalizing Social Network Services Data – Deep Content Analysis to Comprehend Brand Presence*	To operationalize SNS textual data to materialize the role of content and content producer in interactions observed	Adopting more nuanced approaches to utilize dynamic data sources and discover the societal interaction in a major component of the innovation ecosystem known as social network services (SNSs). The findings demonstrated how different categories of content generators and negativity bias is driving the sentimentality of micro communications on SNSs.
IV: *Efficiency Assessment of the Social Capital Capacity on Entrepreneurial Activity: A Perspective Driven from Social Media*	Incorporate SNS data to represent social capital in entrepreneurially-oriented activities among other metrics to assess the efficiency of economies.	Demonstrating the importance of including SNS data in innovation measurement practices. Capturing and incorporating the accurate social capital from SNSs interactions which can reveal undermining capacities in economies which were presented by Data Envelopment Analysis (DEA) efficiency assessment ratio.

V: *Crowd Intelligence Participation in Digital Ecosystem: Systematic Process for Driving Insight from Social Network Services Data*	To examine the escalation of a digital ecosystem by observing how a mass intelligence crowd participates in an SNS platform during disastrous societal challenges	Showing the essential role of SNS platforms in digital ecosystems. A pattern of discussion around coherent topics evolved out of the continued participation of users in the platform causing the formation of coherent topics.

5 CONCLUSIONS

This chapter presents the overall conclusions of this thesis by summarizing the findings and their contribution to this thesis, or their answering of the research questions. Theoretical contributions will be discussed separately, and applications to managerial practice will be outlined. Additionally, the limitations of this research and future research directions are discussed to prompt further exploration into this important topic.

5.1 Addressing the research questions

This thesis focused on the advancement of the ecosystem approach to innovation studies by introducing social network services (SNSs) as a new data source. With the advent of social media platforms, or SNSs, and the massive number of discussions represented in them, it has become necessary to include these valuable data sources, yet the operationalization of user-generated content in SNSs as a major component of the innovation ecosystem has not been effectively and efficiently studied. While the literature on the emerging ecosystem approach towards innovation lays the ground work for undertaking novel big data sources, there is a need to consider the inclusion of SNSs content. Accordingly, the main research objective of the thesis was formulated as *Positioning social network services within the ecosystem approach to innovation.*

The theoretical background of the thesis looked at the measurement and evaluation of innovation from an evolutionary economics and innovation management perspective. Innovation is increasingly recognized as a distributed and collective process that involves a variety of components and their interaction. As a result, innovation performance has been captured using various perspectives, methods and, indicators. The development of innovation indicators over the years clearly indicates a shift towards the recognition of intangible, or soft, aspects of innovation. Applying systems and network science to this concept has triggered an important evolution in innovation evaluation studies, which has resulted in a shift in focus from linear to nonlinear approaches. This evolving process of innovation evaluation is also represented in the shift away from tangible indicators, such as scientific publications and patents, to more intangible indicators, like human capital, social capital, and brand presence. The importance of intangible aspects in these studies is apparent in their inclusion as an additional helix to the traditional triple helix of "university-industry-government relations." Despite the discussion and recognition of intangibles, especially social capital, in various studies, measuring these variables is a complex and difficult process. The thesis first research question "*What is the role of SNSs as one of the multitude of components in the innovation ecosystem?"* set out to investigate the value of social network services when an ecosystem approach to innovation studies is adopted. The first task of this thesis was to understand the growing literature of the "innovation ecosystem." In Publication I, innovation ecosystem literature was observed; by utilizing bibliometric data collected from Web of Science, an indexing database of scientific articles, the major contributors to the debate were identified. The innovation ecosystem as an emerging concept was benchmarked against the closely-related concept

of the innovation system and the analysis also took into account the author's collaboration network and any major topical discussion. The results of the first publication distinguished the innovation ecosystem as an emerging concept with different contributions to the scientific debate on a holistic view of innovation. Further elaboration of this concept took place in Publication II, where an attempt to perform an in-depth analysis of the innovation ecosystem literature by exploring its major attributes was made, resulting in an acceptance of the ecosystem approach of the literature. Publication II establishes the system attributes (Components, Interaction, Dynamism, Environment and Goal) of an ecosystem by revisiting General System Theory, which then called for elaboration of these attributes within the innovation ecosystem literature. The findings in Publication II revealed that the dynamism and goal aspects were not been sufficiently discussed in the innovation ecosystem literature. The definition of dynamism, according to General System Theory, refers to continuous change in a system suggesting that attention should be given to recent live-stream data sources. Social media, or social network services (SNSs), facilitate the creation and sharing of information via virtual communities and/or networks, which can have instant social implication on account of their wide user base. Consecutively, the next step was to use SNS data to address the second research question: *How can SNSs be utilized and materialize as a component in the innovation ecosystem?*

With the advent of mass communication hosted by social network services (SNS) came a novel opportunity to utilize publicly-available data sources to speak to one of the less tangible but more important components of innovation evaluation practices—social capital. Theoretically, the notion of social capital is essential to evaluating innovation or sketching out innovation ecosystems. A practical approach toward utilizing SNS data was taken at the micro-level in Publication III and at the meso-level in Publication IV. Publication III is an attempt to operationalize SNS data by leveraging machine learning applications and natural language processing to gain greater insights. The case study performed in Publication III discovered a relationship between the type of content producers in an SNS (in our case Twitter) and the category that content fits into and the level of interaction their content receives. To exemplify this relationship, an aggregate view of SNSs data was challenged by incorporating profiles of content producers, as these explained the variance in the polarity of content sentiment. The results imposed different carrying capacities (negative and positive) on highly impactful content, in this case: tweets, as the content under investigation. The results revealed that the weight and value of messages with negative polarity far outweighed those with positive and neutral ones. This negative bias can perhaps be explained by the fact that content was generated by both personal and professional user categories.

For a meso-level approach, Publication IV used countries as a level of analysis with the intent to incorporate SNS data for an efficient analysis of the social-economical aspects of the countries studied. Twitter was selected as the SNS for this study as it provided data appropriate for the context of the study. The startup ecosystem of each country was targeted to capture entrepreneurial-related discussion on the platform. The practice of

utilizing SNS data for this study aimed to capture entrepreneurial-oriented activities which the resulting metric consolidated to evaluate economy efficiency.

The last publication of this thesis argued that SNSs played a major role in current digital platforms and took an explorative look at the discussion surrounding a specific disaster situation hosted on an SNS. This study advanced the second research question of this thesis by automatizing much of the analytical process for SNS data. The SNSs data analytical process was designed in a cloud-based system to perform an independent task in one accessible environment. Multiple sets of processes, such as text cleaning, language detection, translation and text categorization were applied in order to detect the major discussion points around the topic on the SNSs. An analysis of periodic topical distance in major discussions on SNSs identified a pattern of collective knowledge formation and dissemination. This systematic methodological process enabled SNS data to provide insightful information both efficiently and effectively and opened a possible avenue for the inclusion of an SNS component in digital platform studies.

Overall, all these publications supported the main objective of the thesis by tackling the two major research questions relating to the conceptual positioning and practical operationalization of SNS data when adopting an ecosystem approach to innovation studies. Not only that, but the application of advanced textual analytical methods helped the researchers better understand, recognize, and support the critical role of SNS data in capacity-building activities.

5.2 Theoretical implications

The theoretical contributions of this dissertation can be categorized into two main areas. The theoretical grounding was built on the literature around an ecosystem approach to innovation studies, with particular attention given to the role of social network services. More specifically, analytical approaches to social network services data were considered to gain insight into the role of this data in the ecosystem. This dissertation contributed to the conceptualization of intangibles within innovation measurement regimes by focusing on social network services as a medium for societal discussion. Earlier work by Phillips (2006) and Carayannis and Campbell (2009) also call for going beyond the technological aspects of the innovation ecosystem to gain insight from soft elements, such as: culture, organization, and institution in order to enrich the dimensions of the innovation ecosystem.

First, this thesis' contribution adds to the conceptualization of the innovation ecosystem, which remains ambiguous despite earlier attempts at conceptualization (Oh *et al.*, 2016). The ecosystem approach to innovation studies was benchmarked with a systems approach to innovation studies which overlaps with the ecosystem approach in many ways. That being said, the bibliometric and content analysis of the two bodies of literature helped to distinguish the innovation ecosystem as its own domain and allow for its dissemination. This knowledge is quite helpful for members of the scientific community interested in the major topical contributions of the innovation ecosystem and how it differs from a systems

approach to innovation. Furthermore, searching through the growing body of published works on the innovation ecosystem, an attempt was made to observe the general system attributes within said literature. The investigation of GST in the innovation ecosystem literature suggests an elaboration of two system attributes: dynamism and goal. A detailed analysis of an ecosystem approach to innovation studies will enable future researchers to comprehend the growing body of knowledge and its value added to new contributions to the literature. The conceptualization of the innovation ecosystem was not only developed but also applied empirically in this dissertation. Detailed attention was paid in three separate studies (Publications III, IV, and V) to social network services data.

Secondly, this thesis demonstrated the use of social network services data at the micro- and meso-level. Publication III was the micro-level perspective that was gleaned from SNSs discussion regarding flagship product of sample of five companies. By exploring the sentimentality of Twitter content with regards to the brands under investigation, the "negativity bias" theory was most interestingly confirmed. This negative bias, also known as the negativity effect theory, states that, even of equal intensity, things of a more negative nature (e.g. unpleasant thoughts, emotions or social interactions; traumatic events) have a greater effect on a person's psychological state and processes than those that are neutral or positive (Lewicka, Czapinski and Peeters, 1992; Baumeister *et al.*, 2001; Rozin and Royzman, 2001). This initial confirmation of the original theory by the data was promising as an example of the capability of data to answer new questions. Earlier literature observed how SNSs might impact interpersonal relationships and, ultimately, the effects of social capital (Utz and Muscanell, 2015). Nonetheless, an analytical approach towards SNS data exploited the network structure among people as well as properties, such as strong and weak ties to social capital (Binder & Sutcliffe 2014; Antheunis et al. 2015; Nicole B. Ellison et al. 2006). Consequently, in this research significant weight was placed on investigating the content in SNSs, whereas the network structure served as a connection between the active entities. Therefore, this publication developed and tested a hypothesis to better understand the role both content producer and type of content played in regards to the sentimentality of the public viewer. The findings in Publication III also contributed to the social media research literature by introducing the variability within data from SNS content. In reference to the causal-chain framework of social media research proposed by Ngai et al. (2015), the spectrum of mediators and moderators was expanded when this research introduced the layers embedded in SNS content (type of generating user and type of content) into the process of moving from antecedents towards outcomes.

5.3 Managerial implications

Innovation ecosystem research has only been performed for two decades, so too little is still known about how firms can fully leverage the added values to enhance innovation performance. However, this thesis attempted to not only theoretically argue for the importance of social network services in the innovation ecosystem but also empirically link its methodological practices for better comprehension and employment of SNS data.

Given the importance of SNSs in understanding the innovation ecosystem, this section will offer applicable and emerging implications for practicing managers.

From a managerial perspective, it is no longer sufficient to treat innovation as a linear process where resources are channeled at one end and new products and/or processes emerge from that end. The innovation measurement and evaluation framework was expanded on to recognize soft aspects of innovation (intangibles) for an accurate understanding of this concept. An ecosystem approach to innovation enables an active study of innovation by considering a diverse set of components and complementarities and embedding interactions among their networks. Mass communication platforms such as SNSs allows for an easy adjustment to and membership in the collaborative network and a transaction of public opinion in ecosystems. The findings of this dissertation imply that social media (here: the SNS) is a component which should be considered when assessing innovation. SNSs are instrumental to businesses across the globe and can now spread a brand message to a wider audience more than ever before. Companies maintain their relationships with public and media over SNSs; and, therefore, SNSs are an important complimentary within a company's innovation ecosystem. The techniques and procedures adopted for conducting the case studies in Publication III can be used by both academics and practitioners to measure a firm's brand presence and marketing of innovation. The process for utilizing SNS data involves targeting relevant discussions, retrieving and preparing data, and constructing and implementing a machine learning model. Moreover, with automation and classification employed through machine learning models, additional features generated from the SNSs data was beneficial in many ways. The process for communicating the process of data processing and analysis and information retrieval and discovery has been included in the publications for replication purposes.

The findings of this thesis demonstrate that the type of profile and type of content that the profile generates in SNSs is perceived differently by users. This implies that marketing managers should consider the role of profiles and their purpose for generating content as their influence on a general audience can vary significantly. This implies that units, such as customer services and advertisements in SNSs, need to be adaptive to be able to tailor to the various needs of their customers. The evidence presented in these cases can serve as the foundation for the ways that SNS data can be operationalized. This viewpoint arose from the realization that SNS data generated novel views and numerous suggestions for the better gauging and operationalizing of data for insightful information, thereby boosting the efficiency and effectiveness of SNS data use (see, for example, Publications III, IV and V). These new insights into the handling of SNS data would benefit businesses who research user profiles before marketing or promoting their products and services. This would give them a competitive edge and, simultaneously highlight the importance of SNSs as a necessary component of the ecosystem approach to analyzing a phenomenon.

Any conclusions for managerial consideration based on this thesis must also be carefully scrutinized for the possibility of alternative explanations. The broad nature of the study

of the innovation ecosystem can extend to a number of entities, and the amount of information that is captured and presented can often be overwhelming to the end-user. As was acknowledged by Basole et al. (2015), the challenge of how the study and conceptualization of a particular innovation ecosystem and its entities is carried out often relates to the context of the problem.

5.4 Assessment of the research

In scientific research, the principles of validity and reliability are the fundamental cornerstones of the scientific method (Creswell, 2014). For a study to be valid, it needs both to address the questions under investigation and use appropriate methods and design to collect and analyze data (Yin, 2013). In other words, the integrity of a study rests on whether its conclusions are valid (Creswell, 2014), with validity referring to the capturing of truth with no regard to outside influences or personal preferences (Saunders *et al.*, 2012). On the other hand, reliability discusses the consistency or repeatability of research measures. The reliability aspect of a research project can be improved in several ways. One way is to carefully document the different phases of the research process as they proceed. According to Yin (2013), this documentation can improve reliability in a case study as it enables the same practices to be performed all over again.

In this research, the Design Science Research Methodology (DSRM) was considered and the principles by Peffers (2007) were employed to carry out research that met the study objective. By definition, the DSR process is to design a purposeful artifact to address a relevant problem (Hevner *et al.*, 2004; Peffers, Tuunanen and Rothenberger, 2007). Artifact construction should be accomplished in a transparent and replicable way, demonstrating practical feasibility as well as methodological validity (Hevner *et al.*, 2004). In order for other researchers to be able to reconstruct the design experiment, a description of the procedure for utilizing the artifact (e.g., test scripts, scenarios, flowcharts) and the way to collect and analyze data (e.g., step-by-step description, validation protocols) are required (Mettler, Eurich and Winter, 2014).

In this study, the aim was to position social network services (SNSs) within the ecosystem approach towards innovation studies. This approach of utilizing and materializing SNSs as an attribute in the innovation ecosystem involved the novel use of SNS data analytics. The utility of the artifact is examined in various case studies (Publications III, IV and V). The literature necessary for conceptualizing SNSs as attributes of the innovation ecosystem was systematically collected and analyzed (section 3.1). Furthermore, to operationalize SNSs, data was obtained from the popular application known as Twitter and systematically collected and analyzed (Section 3.2). Internal validity was maintained in all of the publications where the artifact was in practice. Testing for the validity of the construct regarding SNS data processing optimally followed the exact same procedures as that of statistical models. A relevant accuracy measure for the sentiment classifier in Publication III was calculated. This indicates the acceptable accuracy rate for implementing the Naive Bayes classifier to detecting tweet sentiment. For example,

accuracy was reported as the proportion of true results (both true positives and true negatives) in the total number of cases examined in Publication III.

This study adopted two major methodological approaches which were carried out during this research: systematic review and bibliometric analysis of literature and social network services data collection and analytics have been documented in detail. Eventually, documenting the research process in more detail will increase the reliability of this research as it will make the process replicable for other researchers. Presently, all of the sub-studies of this research are published in the form of separate research articles; the multiple case study as research project reports.

5.5 Limitations and suggestions for future research

Due to the emerging nature of the research under investigation in this thesis, the challenges and recommendations are more relevant than ever for both academics and business professionals. The objective of this thesis was to raise awareness of the roles of social network services in the innovation ecosystem context. This dissertation shows that it is crucial to place the soft aspects and intangibilities into consideration while studying the innovation ecosystem. With regards to using social network services data to better understand the innovation ecosystem, this thesis provided preliminary insight into the topic. SNSs are evolving platforms—changing due to user expectations while simultaneously causing user behavior to change. In short, we can conclude that the way that SNSs were used last year changed the way that they are being used this year. This dynamic feedback system requires an evolving approach to its measurement. Moreover, this evolving nature can also be applied to characteristics of the innovation ecosystem on account of its ability to adapt and evolve (Adomavicius *et al.*, 2006; Basole, 2009).

The availability of data from SNS platforms can be considered one of the major limitations for doing research on SNSs data. While some platforms offer public Application Programming Interface (API)s to retrieve at least a sample set of historical and livestream data, this did not hold for other major SNSs platforms. Users exploit different SNSs for different purposes; therefore, it is necessary to observe other SNSs platforms. In all of the case studies, Twitter data was used because its structure and availability through an open access API made data collection feasible. On the other hand, the structure and policies of other platforms, such as Facebook, make the possibility of utilizing data from this platform impossible for research purposes. Any technical and organizational measures were implemented according to General Data Protection Regulation (GDPR) (EU) 2016/679 during the course of data handling and drafting of these studies.

The skepticism that comes with using new ways and methods of measuring has always been a necessary evil for academics and practitioners, but over time they usually become part of the common base of understanding. As an example, Martin and Irvine's (1984) early work on the use of bibliometrics as an indicator was not appreciated by governments and scientists when it was first released (Brusoni, Prencipe and Salter, 1998; Martin,

1999). However, after only a decade, bibliometrics became one of the official measurement tools used by governments to assess science and research quality.

In a technologically- and digitally-mediated world where various levels of connectivity and interactions can be envisioned for the stakeholders of an ecosystem, it is necessary to be able to distinguish an innovation ecosystem. This creates the need for a new set of skills for diagnosing and designing ecosystems and determining how stakeholders can capture value by monetizing the result and opportunities (Parker, Van Alstyne and Jiang, 2016; Helfat and Raubitschek, 2018). The principle focus of this thesis was gain a greater understanding of how social network services data can inform the study of the innovation ecosystem. The discussion of objectives, methods, and findings in this research provided an intensive review of the importance of SNSs and ways to incorporate their data. The effect of SNSs has been significant and should not be ignored. The direction for future research should help to discover both the positive and negative sides of SNSs in our individual lives, societies, and economies. With all the interest in SNSs, the extending body of research conducted on SNSs will intensify in the future, which should reveal an understanding of the true value of SNSs.

REFERENCES

Acosta, D. M. (2014) 'Tweet Up? Examining Twitter's Impact on Social Capital and Digital Citizenship in Higher Education', *About Campus*, 18(6), pp. 10–17. doi: 10.1002/abc.21139.

Adams, R. *et al.* (2006) 'Innovation management measurement: A review', *International Journal of Management Reviews*, 8(1), pp. 21–47. doi: 10.1111/j.1468-2370.2006.00119.x.

Adams, R., Bessant, J. and Phelps, R. (2006) 'Innovation management measurement: A review', *International Journal of Management Reviews*, 8(1), pp. 21–47. doi: 10.1111/j.1468-2370.2006.00119.x.

Adedoyin-Olowe, M., Gaber, M. M. and Stahl, F. (2014) 'A Survey of Data Mining Techniques for Social Network Analysis', *International Journal of Research in Computer Engineering and Electronics*, 3(6), pp. 1–8. doi: 10.1080/00131881.2016.1220810.

Adner, R. (2006) 'Match Your Innovation Strategy to Your Innovation Ecosystem Match Your Innovation Strategy to Your Innovation Ecosystem', *Harvard business review*.

Adner, R. (2017) 'Ecosystem as Structure: An Actionable Construct for Strategy', *Journal of Management*, 43(1), pp. 39–58. doi: 10.1177/0149206316678451.

Adner, R. and Kapoor, R. (2010) 'Value creation in innovation ecosystems: how the structure of technological interdependence affects firm performance in new technology generations', *Strategic Management Journal*, 31(3), pp. 306–333. doi: 10.1002/smj.821.

Adner, R. and Kapoor, R. (2016) 'Innovation ecosystems and the pace of substitution: Re-examining technology S-curves', *Strategic Management Journal*, 37(4), pp. 625–648. doi: 10.1002/smj.2363.

Adomavicius, G. *et al.* (2006) 'Understanding Patterns of Technology Evolution : An Ecosystem Perspective', *Business*, 0(C), pp. 1–10.

Albino, V., Berardi, U. and Dangelico, R. M. (2015) 'Smart cities: Definitions, dimensions, performance, and initiatives', *Journal of Urban Technology*, 22(1), pp. 1–19. doi: 10.1080/10630732.2014.942092.

Alexa (2018) *The top 500 sites on the web*. Available at: https://www.alexa.com/topsites (Accessed: 15 March 2018).

Allen, P. M., Maguire, S. and McKelvey, B. (2011) 'The SAGE Handbook of Complexity and Management', pp. 79–93.

Amichai-Hamburger, Y., Kingsbury, M. and Schneider, B. H. (2013) 'Friendship: An old concept with a new meaning?', *Computers in Human Behavior*, 29(1), pp. 33–39.

doi: 10.1016/j.chb.2012.05.025.

Anderson, P. (1999) 'Complexity theory and organizational science', *Organizational science*, 10(3), pp. 216–232.

Andriani, P. (2011) 'Complexity and innovation', in *The SAGE handbook of complexity and management*. Sage Publications, pp. 454–470.

Anggarwal, C. C. (2011) *Social Network Data Analytics*. Edited by C. C. Aggarwal. Boston, MA: Springer US. doi: 10.1007/978-1-4419-8462-3.

Anne, H., David, T. and Joan Ernst, V. A. (2006) 'Management as a Design Science Mindful of Art and Surprise', *Journal of Management Inquiry*, 15(4), pp. 413–424. doi: 10.1177/1056492606295900.

Antheunis, M., Abeele, M. and Kanters, S. (2015) 'The Impact of Facebook Use on Micro-Level Social Capital: A Synthesis', *Societies*, 5(2), pp. 399–419. doi: 10.3390/soc5020399.

Archer, Geoff. Van Woensel, L., Van Woensel, L. and Archer, G. (2015) *Ten technologies which could change our lives: Potential impacts and policy implications*. doi: 10.2861/803452.

Archibugi, D. *et al.* (1994) 'Evaluation of the Community Innovation Survey (CIS) - Phase 1'.

Armbruster, H. *et al.* (2008) 'Organizational innovation: The challenge of measuring non-technical innovation in large-scale surveys', *Technovation*, 28(10), pp. 644–657. doi: 10.1016/j.technovation.2008.03.003.

Arora, S. K. (2016) *Social media and innovation ecosystems*. Georgia Tech. Available at: https://smartech.gatech.edu/handle/1853/54929.

Arthur, W. B. (1999) 'Complexity and the Economy', *Science*, 284(5411), pp. 107–109. doi: 10.1126/science.284.5411.107.

Autio, E. and Llewellyn, D. W. T. (2014) 'Innovation ecosystems: Implications for innovation management', *The Oxford handbook of innovation management*, (January), pp. 204–228. doi: 10.1093/oxfordhb/9780199694945.013.012.

Avenier, M. J. (2010) 'Shaping a constructivist view of organizational design science', *Organization Studies*, 31(9–10), pp. 1229–1255. doi: 10.1177/0170840610374395.

Banker, R. D., Charnes, A. and Cooper, W. W. (1984) 'Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis', *Management Science*, 30(9), pp. 1078–1092. doi: 10.1287/mnsc.30.9.1078.

Barn, B., Barat, S. and Clark, T. (2017) 'Conducting Systematic Literature Reviews and Systematic Mapping Studies', in *Proceedings of the 10th Innovations in Software Engineering Conference on - ISEC '17*. New York, New York, USA: ACM Press, pp.

212–213. doi: 10.1145/3021460.3021489.

Basole, R. C. (2009) 'Structural Analysis and Visualization of Ecosystems: A Study of Mobile Device Platforms', in *AMCIS*.

Basole, R. C. *et al.* (2015) 'Understanding business ecosystem dynamics: A data-driven approach', *ACM Transactions on Management Information Systems*, 6(2). doi: 10.1145/2724730.

Basole, R. C. *et al.* (2016) 'Visual decision support for business ecosystem analysis', *Expert Systems with Applications*. Elsevier Ltd, 65, pp. 271–282. doi: 10.1016/j.eswa.2016.08.041.

Baumeister, R. F. *et al.* (2001) 'Bad is stronger than good.', *Review of General Psychology*, 5(4), pp. 323–370. doi: 10.1037/1089-2680.5.4.323.

Baumeister, R. F. and Leary., M. R. (1997) 'Writing narrative literature reviews', *Review of General Psychology*, 24(4), pp. 230–235. doi: 1089-2680/97153.00.

Becker, G. S. (1974) 'Human capital: a theoretical and empirical analysis, with special reference to education', p. 390. Available at: https://books.google.com.ar/books/about/Human_Capital.html?id=9t69iICmrZ0C&redir_esc=y.

Belhocine, N. (2008) 'Treating Intangible Inputs as Investment Goods : The Impact on Canadian GDP', *Growth (Lakeland)*, (1215).

Bergek, A. *et al.* (2008) 'Functions in Innovation System Approaches', *Technological Forecasting and Social Change*, 79(4), pp. 413–432. doi: 10.1016/j.techfore.2006.03.002.

Bijker, W. (1995) 'Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change', *Inside Technology*, p. 390. doi: 10.2307/2077312.

Binder, J. and Sutcliffe, A. (2014) 'The Best of Both Worlds? Online Ties and the Alternating Use of Social Network Sites in the Context of Migration', *Societies*, 4(4), pp. 753–769. doi: 10.3390/soc4040753.

Blair, M. M. and Wallman, S. M. H. (2000) 'Unseen Wealth: Report of the Brookings Task Force on Intangibles', p. 140. Available at: http://books.google.pl/books?id=jpbFnUPi7L8C.

Blei and Blei, D. M. (2003) 'Latent dirichlet allocation', *Journal of machine learning research*, 3(Jan).

Blitzer, J. *et al.* (2007) 'Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification', *Annual Meeting-Association for Computational Linguistics*, 45, p. 440. doi: 10.1109/IRPS.2011.5784441.

Bollen, J., Mao, H. and Zeng, X. (2011) 'Twitter mood predicts the stock market',

Journal of Computational Science, 2(1), pp. 1–8. doi: 10.1016/j.jocs.2010.12.007.

Bontis, N. (2001) 'Assessing knowledge assets: a review of the models used to measure intellectual capital', *International Journal of Management Reviews*, 3(1), pp. 41–60. doi: 10.1111/1468-2370.00053.

Boyd, D. m. and Ellison, N. B. (2007) 'Social Network Sites: Definition, History, and Scholarship', *Journal of Computer-Mediated Communication*, 13(1), pp. 210–230. doi: 10.1111/j.1083-6101.2007.00393.x.

Bramwell, A., Hepburn, N. and Wolfe, D. A. (2012) 'Growing Innovation Ecosystems : University-Industry Knowledge Transfer and Regional Economic Development in Canada- Final Report to the Social Sciences and Humanities Research Council of Canada', p. 62. Available at: http://sites.utoronto.ca/progris/presentations/pdfdoc/2012/Growing Innovation Ecosystems15MY12.pdf.

Breuer, A. (2011) 'Democracy Promotion in the Age of Social Media: Risks and Opportunities', *SSRN Electronic Journal*. doi: 10.2139/ssrn.2127198.

Brusoni, S., Prencipe, A. and Salter, A. (1998) *Mapping and Measuring Innovation in Project-based Firms*, *COPS*. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.118.9033&rep=rep1&type=pdf.

Burke, M., Kraut, R. and Marlow, C. (2011) 'Social capital on Facebook: Differentiating uses and users', *CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.*, pp. 571–580. doi: 10.1145/1978942.1979023.

Canibano, L. *et al.* (1999) 'Measuring intangibles to understand and improve innovation management', *OECD Symposium on Measuring and Reporting of Intellectual Capital: Experience, Issues and Prospects*, pp. 1–24.

Capra, F. (1997) 'The web of life', *New York: Doubleday-Anchor Book*.

Carayannis, E. G. and Campbell, D. F. J. (2009) '"Mode 3" and "Quadruple Helix": toward a 21st century fractal innovation ecosystem', *International Journal of Technology Management*, 46(3/4), p. 201. doi: 10.1504/IJTM.2009.023374.

Carayannis, E. G. and Economy, R. R.-J. of the K. (2014) 'The quadruple/quintuple innovation helixes and smart specialisation strategies for sustainable and inclusive growth in Europe and beyond', *Springer*. Available at: http://link.springer.com/article/10.1007/s13132-014-0185-8.

Carlsson, B. and Stank (1991) 'On the nature, function and composition of technological systems', *Journal of Evolutionary Economics*, 1, pp. 93–118.

Carmichael, D., Archibald, J. and Lund, G. (2015) 'Social Capital Theory in Social Media Research', *SSRN Electronic Journal*. doi: 10.2139/ssrn.2612872.

Castellacci, F. (2014) *The Innovation Union in Europe: A Socio-Economic Perspective on EU Integration*, *Regional Studies*. doi: 10.1080/00343404.2014.886477.

Charnes, A., Cooper, W. W. and Rhodes, E. (1978) 'Measuring the efficiency of decision making units', *European Journal of Operational Research*, 2(6), pp. 429–444. doi: 10.1016/0377-2217(78)90138-8.

Chen, G. K. C. (1975) 'What is the Systems Approach?', *Interfaces*, 6(1), pp. 32–37. doi: 10.1287/inte.6.1.32.

Claridge, T. (2004) *Social Capital and Natural Resource Management: An important role for social capital?*, *Natural and Rural Systems Management*. University of Queensland. Available at: https://www.socialcapitalresearch.com/wp-content/uploads/2013/01/Social-Capital-and-NRM.pdf.

Coleman, J. (2000) 'Social Capital in the Creation of Human Capital', *Knowledge and Social Capital*, pp. 17–41. doi: 10.1016/B978-0-7506-7222-1.50005-2.

Cooper, G. P. *et al.* (2015) 'Twitter as a potential disaster risk reduction tool. part i: Introduction, terminology, research and operational applications', *PLoS Currents*, 7(DISASTERS). doi: 10.1371/currents.dis.a7657429d6f25f02bb5253e551015f0f.

Cooper, H. (2003) 'Psychological Bulletin: Editorial.', *Psychological Bulletin*, 129(1), pp. 3–9. doi: 10.1037/0033-2909.129.1.3.

Corbett, A. C. (2007) 'Learning asymmetries and the discovery of entrepreneurial opportunities', *Journal of Business Venturing*, 22(1), pp. 97–118. doi: 10.1016/j.jbusvent.2005.10.001.

Cordero, R. (1990) 'The measurement of innovation performance in the firm: An overview', *Research Policy*, 19(2), pp. 185–192. doi: 10.1016/0048-7333(90)90048-B.

Corrado, C., Hulten, C. and Sichel, D. (2006) *Intangible Capital and Economic Growth*. Cambridge, MA. doi: 10.3386/w11948.

Creswell, J. W. (2014) 'Research design : qualitative, quantitative, and mixed methods approaches', p. 273. doi: http://dx.doi.org/10.1016/j.math.2010.09.003.

Dahlstrand, Å. L. and Stevenson, L. (2010) 'Innovative entrepreneurship policy: linking innovation and entrepreneurship in a European context', *Annals of Innovation & Entrepreneurship*, 1, pp. 1–15. doi: 10.3402/aie.v1i1.5602.

Das, S. R. and Chen, M. Y. (2007) 'Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web', *Management Science*, 53(9), pp. 1375–1388. doi: 10.1287/mnsc.1070.0704.

Dedehayir, O., Mäkinen, S. J. and Roland Ortt, J. (2016) 'Roles during innovation ecosystem genesis: A literature review', *Technological Forecasting and Social Change*. Elsevier Inc., pp. 1–12. doi: 10.1016/j.techfore.2016.11.028.

Drucker, P. (2014) *Innovation and Entrepreneurship*. Routledge.

Durlauf, S. and Fafchamps, M. (2004) *Social Capital*. Cambridge, MA. doi: 10.3386/w10485.

Edquist, C. (1997) *Systems of Innovation: Technologies, Institutions and Organizations*. Cassell Academic.

Eichstaedt, J. C. *et al.* (2015) 'Psychological Language on Twitter Predicts County-Level Heart Disease Mortality', *Psychological Science*, 26(2), pp. 159–169. doi: 10.1177/0956797614557867.

Ellison, N., Steinfield, C. and Lampe, C. (2007) 'The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites', *Jcmc*, 12(4), pp. 1143–1168. doi: 10.1111/j.1083-6101.2007.00367.x.

Ellison, N., Steinfield, C. and Lampe, C. (2006) 'Spatially Bounded Online Social Networks and Social Capital: The Role of Facebook', *Annual Conference of the International Communication Association*, p. 0. doi: 10.1.1.85.5541.

Eom, Y. H. *et al.* (2015) 'Twitter-based analysis of the dynamics of collective attention to political parties', *PLoS ONE*, 10(7). doi: 10.1371/journal.pone.0131184.

Etzkowitz, H. and Leydesdorff, L. (2000) 'The dynamics of innovation: from National Systems and Mode 2'' to a Triple Helix of university–industry–government relations', *Research Policy*, 29, pp. 109–123. doi: 10.1016/S0048-7333(99)00055-4.

European Commission (2004) *European Innovation Scoreboard Methodology Report*.

European Commission (2013) 'Research and Innovation performance in EU Member States and Associated countries', p. 334. doi: 10.2777/82363.

European Union (2015) *Growing A Digital Social Innovation Ecosystem For Europe : DSI Final Report*. doi: 10.2759/448169.

European Union (2016) *Regional innovation ecosystems CoR guide : learning from the EU's cities and regions*. Available at: https://publications.europa.eu/en/publication-detail/-/publication/6a43bcbb-85a9-43fc-afa3-db58c42f4730/language-en.

European Union (2017) *Place-Based Innovation Ecosystems: Espoo Innovation Garden and Aalto University (Finland)*. doi: 10.2760/31587.

Evans, P. C. and Basole, R. C. (2016) 'Revealing the API ecosystem and enterprise strategy via visual analytics', *Communications of the ACM*, 59(2), pp. 26–28. doi: 10.1145/2856447.

Fairlie, R. W. and Robb, A. (2007) 'Families, human capital, and small business: Evidence from the characteristics of business owners survey', *Industrial and Labor Relations Review*, 60(2), pp. 225–245. doi: 10.1177/001979390706000204.

Florida, R. (2012) *The Rise of the Creative Class--Revisited: 10th Anniversary Edition--*

Revised and Expanded. Basic Books. Available at: http://www.amazon.com/The-Rise-Creative-Class-Revisited-Edition-Revised/dp/0465029930.

Foray, D. (2004) *Economics of Knowledge*. MIT Press.

Freeman, C. and Luc, S. (1997) *The economics of industrial innovation*. Psychology Press.

Frenkel, A. *et al.* (2001) 'Supporting Information Access Rights and Visibility Levels in Virtual Enterprises', in *E-Business and Virtual Enterprises*. Boston, MA: Springer US, pp. 177–192. doi: 10.1007/978-0-387-35399-9_17.

Fukuyama, F. (1995) *Trust: The Social Virtues and the Creation of Prosperity*. Free Press.

Gartner (2017) *Social Analytics*. Available at: https://www.gartner.com/it-glossary/social-analytics (Accessed: 5 March 2018).

Gawer, A. (2014) 'Bridging differing perspectives on technological platforms: Toward an integrative framework', *Research Policy*. Elsevier B.V., 43(7), pp. 1239–1249. doi: 10.1016/j.respol.2014.03.006.

Geissdoerfer, M. *et al.* (2017) 'The Circular Economy – A new sustainability paradigm?', *Journal of Cleaner Production*, 143, pp. 757–768. doi: 10.1016/j.jclepro.2016.12.048.

Gerrish, S. M. (2010) 'A Language-based Approach to Measuring Scholarly Impact', *Computer*, 180(33), pp. 375–382. doi: 10.1002/chin.200533198.

del Giudice, M., Carayannis, E. G. and Della Peruta, M. R. (2012) 'Cross-cultural knowledge management: Fostering innovation and collaboration inside the multicultural enterprise', *Cross-Cultural Knowledge Management: Fostering Innovation and Collaboration Inside the Multicultural Enterprise*, pp. 1–166. doi: 10.1007/978-1-4614-2089-7.

Glänzel, W. (2015) 'Bibliometrics-aided retrieval: where information retrieval meets scientometrics', *Scientometrics*, 102(3), pp. 2215–2222. doi: 10.1007/s11192-014-1480-7.

Go, A., Bhayani, R. and Huang, L. (2009) 'Twitter sentiment classification using distant supervision', *CS224N Project Report, Stanford*. Available at: http://cs.wmich.edu/~tllake/fileshare/TwitterDistantSupervision09.pdf (Accessed: 15 October 2013).

Gomes, L. A. de V. *et al.* (2015) 'Unpacking the innovation ecosystem construct: Evolution, gaps and trends', *Technological Forecasting and Social Change*. Elsevier Inc. doi: 10.1016/j.techfore.2016.11.009.

Goodman, M. (2015) *Importance of Systems Thinking Today*, *appliedsystemsthinking*. Available at: http://www.appliedsystemsthinking.com/importance.html (Accessed: 20

February 2017).

Granovetter, M. S. (1985) 'Economic Action and Social Structure : The Problem of Embeddedness', 91(3), pp. 481–510.

Greiling, D. (2006) 'Performance measurement : a remedy for increasing the efficiency of public services ?', *Internation Journal of Productivity and Performance Measurement*, 55(6), pp. 488–465. doi: 10.1108/17410400610682488.

Griliches, Z. (1990) 'Patent Statistics as Economic Indicators: A Survey.', *Journal of Economic Literature*, 28(4), pp. 1661–1707. doi: 10.1016/S0169-7218(10)02009-5.

Grimaldi, M. *et al.* (2017) 'A systematic literature review on intangible assets and open innovation', *Knowledge Management Research & Practice*, 15(1), pp. 90–100. doi: 10.1057/s41275-016-0041-7.

Grimmer, J. (2009) 'A Bayesian Hierarchical Topic Model for Political Texts : Supplemental Appendix Variational Estimation of the Expressed Agenda Model', *Machine Learning*, 1, pp. 1–8.

Grossman, G. M. and Elhanan, H. (1993) *Innovation and growth in the global economy*, *MIT Press*. MIT Press.

Hage, Jerald; Mote, Jonathon E; Jordan, G. B. (2013) 'Ideas, innovations, and networks: a new policy model based on the evolution of knowledge', *Policy Sciences (Jun 2013)*, 46(2), pp. 199–216. doi: 10.1007/s11077-012-9172-8.

Hajikhani, A. (2017) 'Emergence and dissemination of ecosystem concept in innovation studies : A systematic literature review study', in *Hawaii International Conference on System Sciences (HICSS) 2017*, pp. 1–12. Available at: http://hdl.handle.net/10125/41796.

Hannah, D. P. (2013) 'Puzzles or Pieces: Competition in Nascent System Industries', *Academy of Management Proceedings*, 2013(1), pp. 14988–14988. doi: 10.5465/AMBPP.2013.14988abstract.

Hansen, M. and Roll, K. (2016) 'Social capital and adoption of agronomic practices: Theory and findings', (MiB Working Paper no. 2), pp. 1–26.

Helfat, C. E. and Raubitschek, R. (2018) 'Dynamic and Integrative Capabilities for Profiting From Innovation in Digital Platform-Based Ecosystems', *SSRN Electronic Journal*. doi: 10.2139/ssrn.3122046.

den Hertog, P., Bilderbeek, R. and Maltha, S. (1997) 'Intangibles', *Futures*, 29(1), pp. 33–45. doi: 10.1016/S0016-3287(96)00064-X.

Hevner, A.; Chatterjee, S. (2010) 'Design research in information systems: theory and practice':, *Springer Science & Business Media*, 22.

Hevner, A. R. *et al.* (2004) 'Design science in information systems research', *MIS Q.*,

28(1), pp. 75–105.

von Hippel, E. (2007) 'The Sources of Innovation', *Das Summa Summarum des Management*, pp. 111–120. doi: 10.1007/978-3-8349-9320-5_10.

Hofer, M. and Aubert, V. (2013) 'Perceived bridging and bonding social capital on Twitter: Differentiating between followers and followees', *Computers in Human Behavior*, 29(6), pp. 2134–2142. doi: 10.1016/j.chb.2013.04.038.

Hughes, A. L. and Palen, L. (2009) 'Twitter adoption and use in mass convergence and emergency events', *International Journal of Emergency Management*, 6(3/4), p. 248. doi: 10.1504/IJEM.2009.031564.

Huhtamäki, J. (2016) *Ostinato Process Model for Visual Network Analytics Jukka Huhtamäki Ostinato Process Model for Visual Network Analytics Experiments in Innovation Ecosystems Julkaisu 1425 • Publication 1425 Tampere 2016*.

Huhtamäki, J. and Rubens, N. (2016) 'Exploring Innovation Ecosystems as Networks: Four European Cases', *Proceedings of the Hawaii International Conference on System Sciences HICSS-49: January 5-8, 2016, Grand Hyatt, Kauai*, p. 10. doi: 10.1109/HICSS.2016.560.

Hyrynsalmi, S. *et al.* (2012) 'App store, marketplace, play! an analysis of multi-homing in mobile software ecosystems', *CEUR Workshop Proceedings*, 879, pp. 59–72. doi: 10.1007/s00199-006-0114-6.

Håkansson, H. (1990) 'Technological collaboration in industrial networks', *European Management Journal*, 8(3), pp. 371–379. doi: 10.1016/0263-2373(90)90016-Y.

Iansiti, M. and Levien, R. (2004) 'Keystones and dominators: framing operating and technology strategy in a business ecosystem', *Harvard Business School, Working Paper*, pp. 3–61.

Internet Scociety (2014) *Global Internet Report*. Available at: https://www.internetsociety.org/sites/default/files/Global_Internet_Report_2014.pdf.

Irvine Ben, J. & M. (1984) 'Foresight in science', p. 166.

IT for Change (2013) *ICTs for empowerment and social transformation - A note prepared by IT for Change for ActionAid International*. Available at: https://www.itforchange.net/ICTs_for_empowerment_and_social_transformation-A_note_prepared_by_IT_for_Change_for_ActionAid_International_HTML.

Jacobides, M. G., Cennamo, C. and Gawer, A. (2018) 'Towards a theory of ecosystems', *Strategic Management Journal*, (October 2016), pp. 1–22. doi: 10.1002/smj.2904.

Jacobides, M. G., Knudsen, T. and Augier, M. (2006) 'Benefiting from innovation: Value creation, value appropriation and the role of industry architectures', *Research Policy*, 35(8 SPEC. ISS.), pp. 1200–1221. doi: 10.1016/j.respol.2006.09.005.

Jacobides, M., Veloso, F. and Wolter, C. (2014) *Ripples through the value chain and positional bottlenecks: Innovation and profit evolution in a competitive setting*. London.

Jaeger, P. T. *et al.* (2007) 'Community response grids: E-government, social networks, and effective emergency management', *Telecommunications Policy*, 31(10–11), pp. 592–604. doi: 10.1016/j.telpol.2007.07.008.

Jarboe, K. P. and Ellis, I. (2010) 'Intangible Assets - Innovative Financing for Innovation', *Issues in Science and Technology*. doi: 10.2307/43315141.

Jarboe, K. P. and Furrow, R. (2008) 'Intangible Asset Monetization: The Promise and the Reality', *Information, Innovation, Intangible Economy*, Working Pa(April), p. 141.

Jenkins, G. M. and Youle, P. V. (1968) 'A Systems Approach to Management', *OR*, 19, p. 5. doi: 10.2307/3007468.

Jensen, M. B. *et al.* (2007) 'Forms of knowledge and modes of innovation', *Research Policy*, 36(5), pp. 680–693. doi: 10.1016/j.respol.2007.01.006.

Johannessen, J. a (2013) 'Innovation: a systemic perspective - developing a systemic innovation theory', *Kybernetes*, 42(8), pp. 1195–1217. doi: 10.1108/k-04-2013-0069.

Jones, P. (2001) 'Are educated workers really more productive?', *Journal of Development Economics*, 64(1), pp. 57–79. doi: 10.1016/S0304-3878(00)00124-3.

Kaasa, A. (2007) *Effects of Different Dimensions of Social Capital on Innovation: Evidence from Europe at the Regional Level*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=976871.

Kaigo, M. (2012) 'Social media usage during disasters and social capital: Twitter and the Great East Japan earthquake', *Keio Communication Review*, (34), pp. 19–35. Available at: http://www.mediacom.keio.ac.jp/publication/pdf2012/KCR34_02KAIGO.pdf.

Kapoor, R. and Lee, J. M. (2013) 'Coordinating and competing in ecosystems: How organizational forms shape new technology investments', *Strategic Management Journal*, 34(3), pp. 274–296. doi: 10.1002/smj.2010.

Kasanen, A., Lukka, K. and Siitonen, A. (1993) 'The Constructive Approach in Management Accounting Research', *Journal of Management Accounting Research*, 5(June 1991), pp. 243–264. doi: http://dx.doi.org/10.1108/17506200710779521.

Khandekar, S. and Phani, B. V. (2017) 'Innovation and Entrepreneurship Ecosystem at IIT Kanpur: A Journey of Serendipity', in, pp. 5–14. doi: 10.1007/978-981-10-3334-6_2.

Kireyev, K., Palen, L. and Anderson, K. (2009) 'Applications of topics models to analysis of disaster-related twitter data', *NIPS Workshop on Applications for Topic Models: Text and Beyond*, p. Available at: http://www.umiacs.umd.edu/~jbg/nips_tm_workshop/15.pdf.

Kitchenham, B. (2004) 'Procedures for performing systematic reviews', *Keele, UK, Keele University*, 33(TR/SE-0401), p. 28. doi: 10.1.1.122.3308.

Klein, P. G. and Cook, M. L. (2006) 'T.W. Schultz and the human-capital approach to entrepreneurship', *Review of Agricultural Economics*, 28(3), pp. 344–350. doi: 10.1111/j.1467-9353.2006.00297.x.

Klevorick, A. K. *et al.* (1995) 'On the sources and significance of interindustry differences in technological opportunities', *Research Policy*, 24(2), pp. 185–205. doi: 10.1016/0048-7333(93)00762-I.

Knutas, A., Hajikhani, A., Salminen, J., Ikonen, J., Porras, J., (2015) 'Cloud-based Bibliometric Analysis Service for Systematic Mapping Studies', in *Proceedings of the 16th International Conference on Computer Systems and Technologies*, pp. 184–191. doi: 10.1145/2812428.2812442.

Kobayashi, T., Ikeda, K. and Miyata, K. (2006) 'Social capital online: Collective use of the Internet and reciprocity as lubricants of democracy', *Information Communication and Society*, 9(5), pp. 582–611. doi: 10.1080/13691180600965575.

Kolleck, N. (2013) 'Social network analysis in innovation research: using a mixed methods approach to analyze social innovations', *European Journal of Futures Research*, 1(1), p. 25. doi: 10.1007/s40309-013-0025-2.

Koschatzky, K. (2005) 'The Regionalizationof Innovation Policy: New Options for Regional Change?', in Fuchs, G. and Shapira, P. (eds) *Rethinking Regional Innovation and Change: Path Dependency or Regional Breakthrough*. New York: Springer (Economics of Science, Technology and Innovation.

Krlev, G., Bund, E. and Mildenberger, G. (2014) 'Measuring What Matters—Indicators of Social Innovativeness on the National Level', *Information Systems Management*, 31(3), pp. 200–224. doi: 10.1080/10580530.2014.923265.

Kuechler, W. and Vaishnavi, V. (2008) 'On Theory Development in Design Science Research: Anatomy of a Research Project', *European Journal of Information Systems*, 17(5), pp. 489–504.

Kwak, K., Kim, W. and Park, K. (2017) 'Complementary multiplatforms in the growing innovation ecosystem: Evidence from 3D printing technology', *Technological Forecasting and Social Change*. Elsevier, (December 2015), pp. 1–16. doi: 10.1016/j.techfore.2017.06.022.

Lanjouw, J. O. and Schankerman, M. (2012) 'Patent Quality and Research Productivity : Measuring Innovation With Multiple Indicators *', *Society*, 114(495), pp. 441–465.

Legner, C. *et al.* (2017) 'Digitalization: Opportunity and Challenge for the Business and Information Systems Engineering Community', *Business & Information Systems Engineering*, 59(4), pp. 301–308. doi: 10.1007/s12599-017-0484-2.

Lev, B. (2000) *Intangibles: Management, Measurement, and Reporting*. Brookings Institution Press.

Lewicka, M., Czapinski, J. and Peeters, G. (1992) 'Positive-negative asymmetry or "When the heart needs a reason"', *European Journal of Social Psychology*, 22(5), pp. 425–434. doi: 10.1002/ejsp.2420220502.

Leydesdorff, L., Dolfsma, W. and Van Der Panne, G. (2006) 'Measuring the knowledge base of an economy in terms of triple-helix relations among "technology, organization, and territory"', *Research Policy*, 35(2), pp. 181–199. doi: 10.1016/j.respol.2005.09.001.

Leydesdorff, L. and Etzkowitz, H. (1996) 'Emergence of a Triple Helix of university—industry—government relations', *Science and Public Policy*. doi: 10.1093/spp/23.5.279.

Lievrouw, L. A. and Livingstone, S. M. (2002) *Handbook of new media: social shaping and consequences of ICTs*. doi: neue medien.

Lin, N. (2002) 'Social capital: a theory of social structure and action (structural analysis in the social sciences)', *A theory of social structure and action*.

Lindsay, B. R. (2011) 'Social Media and Disasters: Current Uses, Future Options and Policy Considerations.', *Congressional Research Service Reports*, p. 13. Available at: http://fas.org/sgp/crs/homesec/R41987.pdf.

Luce, L. (2012) *Twitter sentiment analysis using Python and NLTK*. Available at: http://www.laurentluce.com/posts/twitter-sentiment-analysis-using-python-and-nltk/.

Ludwig Von Bertalanffy (1972) 'The History and Status of General Systems Theory', *The Academy of Management Journal*, 15(4), pp. 407–426.

Lukka, K. (2003) 'The Constructive Research Approach', *In: Case study research in logistics*, Series B, pp. 83–101.

Lundvall, B. (1992) *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. Pinter Publishers. Available at: https://books.google.fi/books/about/National_Systems_of_Innovation.html?id=B_C3AAAAIAAJ&redir_esc=y.

Lusch, R. F. and Nambisan, S. (2015) 'Service Innovation: A Service-Dominant Logic Perspective', *MIS Quarterly*, 39(1), pp. 155–175. doi: 10.25300/MISQ/2015/39.1.07.

Mairesse, J. and Mohnen, P. (2007) 'A survey of innovation surveys: Taking stock of a growing literature', *... -Banque de France Conference on Innovation*, pp. 1–16. Available at: http://www.cepr.org/meets/wkcn/6/6658/papers/Mairesse_Mohnen.pdf.

Mangold, W. G. and Faulds, D. J. (2009) 'Social media: The new hybrid element of the promotion mix', *Business Horizons*, 52(4), pp. 357–365. doi: 10.1016/j.bushor.2009.03.002.

Mansfield, E. and Griliches, Z. (1984) 'R&D and Innovation: Some Empirical

Findings', pp. 127–148.

March, S. T., Smith, G. F. and Smith, D. (1995) 'Design and natural science research on information technology', *Decision Support Systems*, 15(4), pp. 251–266. doi: 10.1016/0167-9236(94)00041-2.

Marr, B. (2017) *9 Technology Mega Trends That Will Change The World In 2018*, *Forbes*. Available at: https://www.forbes.com/sites/bernardmarr/2017/12/04/9-technology-mega-trends-that-will-change-the-world-in-2018/ (Accessed: 12 August 2018).

Martin, B. R. (1999) 'Technology Foresight in a Rapidly Globalizing Economy', *Forward Thinking: Keys to the Future in Education and Research*, (1991), pp. 1–19. Available at: https://www.sussex.ac.uk/webteam/gateway/file.php?name=Fac-BRM-UNIDO-TF&site=25.

McCain, K. W. (1990) 'Mapping Authors in Intellecual Space: A Technical Overview', *Journal of the American Society for Information Science*, 41(6), pp. 433–443.

Mele, C., Pels, J. and Polese, F. (2010) 'A Brief Review of Systems Theories and Their Managerial Applications', *Service Science*, 2(1–2), pp. 126–135. doi: 10.1287/serv.2.1_2.126.

Mettler, T., Eurich, M. and Winter, R. (2014) 'On the use of experiments in design science research: A proposition of an evaluation framework', *Communications of the Association for Information Systems*, 34(1), pp. 223–240.

Meyer, P. B. and Harper, M. J. (2005) 'Preliminary Estimates of Multifactor Productivity Growth', *Monthly Labor Review*, 128(6), pp. 32–43. Available at: http://stats.bls.gov/opub/mlr/mlrhome.htm%5Cnhttp://search.ebscohost.com/login.aspx?direct=true&db=eoh&AN=0813008&site=ehost-live.

Milbergs, E. and Vonortas, N. (2004) 'Innovation Metrics: Measurement to Insight', *Center for Accelerating Innovation and George Washington University*, p. 7. doi: 10.1108/02580540910943550.

Moore, J. F. (1993) 'Predators and prey: a new ecology of competition.', *Harvard Business Review*, 71(3), pp. 75–86. doi: Article.

Moore, J. F. (1996) 'The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems', *Leadership*, p. 297. doi: 10.1017/CBO9781107415324.004.

Mulas, V., Minges, M. and Applebaum, H. (2016) *Boosting tech innovation ecosystems in cities : a framework for growth and sustainability of urban tech innovation ecosystems*, *Innovations*. Available at: https://www.mitpressjournals.org/doi/abs/10.1162/inov_a_00251.

Muralidharan, S. *et al.* (2011) 'Hope for Haiti: An analysis of Facebook and Twitter usage during the earthquake relief efforts', *Public Relations Review*, 37(2), pp. 175–177. doi: 10.1016/j.pubrev.2011.01.010.

Nair, M. (2011) 'Understanding and Measuring the value of social media', *Wiley periodicals*, pp. 45–51. doi: 10.1002/jcaf.

Nelson, R. R. (1992) 'National innovation systems: A retrospective on a study', *Industrial and Corporate Change*. doi: 10.1093/icc/1.2.347.

Nelson, R. R. and Phelps, E. S. (1966) 'Investment in humans, technological diffusion, and economic growth', *American Economic Review*, 56(2), pp. 69–75. doi: 10.2307/1821269.

NESTA (2009) 'Soft innovation. Towards a more complete picture of innovative change', *Nesta*, (July), pp. 1–112. doi: 10.1093/acprof:oso/9780199572489.001.0001.

Ngai, E. W. T., Tao, S. S. C. and Moon, K. K. L. (2015) 'Social media research: Theories, constructs, and conceptual frameworks', *International Journal of Information Management*. Elsevier Ltd, 35(1), pp. 33–44. doi: 10.1016/j.ijinfomgt.2014.09.004.

Niall, D. (2016) *How technology will change the future of work*, *World Economic Forum*. Available at: https://www.weforum.org/agenda/2016/02/the-future-of-work/ (Accessed: 12 August 2018).

Nicolaisen, J. (2010) 'Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics', *Journal of the American Society for Information Science and Technology*, 61(1), pp. 205–207.

OECD (1996) *The Knowledge-Based Economy*. Paris. Available at: https://www.oecd.org/sti/sci-tech/1913021.pdf.

OECD (2010a) 'Measuring Innovation: A New Perspective', *Perspective*, p. 125. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Measuring+Innovation:+A+New+Perspective#0.

OECD (2010b) *Towards a Measurement Agenda for Innovation*. Available at: https://www.oecd.org/site/innovationstrategy/45392693.pdf.

Oh, D.-S. *et al.* (2016) 'Innovation ecosystems: A critical examination', *Technovation*. doi: 10.1016/j.technovation.2016.02.004.

Olshannikova, E. *et al.* (2017) 'Conceptualizing Big Social Data', *Journal of Big Data*, 4(1), p. 3. doi: 10.1186/s40537-017-0063-x.

Pang, B. and Lee, L. (2008) 'Opinion mining and sentiment analysis', *Foundations and Trends in Information Retrieval*, 2(1), pp. 1–135.

Pang, B., Lee, L. and Vaithyanathan, S. (2002) 'Thumbs up?: sentiment classification using machine learning techniques', *... -02 conference on Empirical methods ...*. Available at: http://dl.acm.org/citation.cfm?id=1118704 (Accessed: 15 October 2013).

Papaioannou, T., Wield, D. V. and Chataway, J. C. (2009) 'Knowledge ecologies and

ecosystems? An empirically grounded reflection on recent developments in innovation systems theory', *Environment and Planning C: Government and Policy*, 27(2), pp. 319–339. doi: 10.1068/c0832.

Parker, G., Van Alstyne, M. W. and Jiang, X. (2016) 'Platform Ecosystems: How Developers Invert the Firm', *Ssrn*. doi: 10.2139/ssrn.2861574.

Patel, P. and Pavitt, K. (1994) 'National innovation systems: Why they are important, and how they might be measured and compared', *Economics of Innovation and New Technology*, 3(1), pp. 77–95. doi: 10.1080/10438599400000004.

Pavitt, K. (1984) 'Sectoral patterns of technical change: Towards a taxonomy and a theory', *Research Policy*, 13(6), pp. 343–373. doi: 10.1016/0048-7333(84)90018-0.

Peffers, K. *et al.* (2006) 'The design science research process: a model for producing and presenting information systems research', *Proceedings of the first international conference on design science research in information systems and technology*.

Peffers, K., Tuunanen, T. and Rothenberger, M. (2007) 'A design science research methodology for information systems research', *Journal of*, 24(3), pp. 45–77. Available at: http://www.tandfonline.com/doi/abs/10.2753/MIS0742-1222240302.

Peltoniemi, M. and Vuori, E. (2004) 'Business ecosystem as the new approach to complex adaptive business environments', *Proceedings of eBusiness Research Forum*, pp. 267–281.

Pénard, T. and Poussing, N. (2010) 'Internet Use and Social Capital: The Strength of Virtual Ties', *Journal of Economic Issues*, 44(3), pp. 569–595. doi: 10.2753/JEI0021-3624440301.

Pentland, A. (2014) *Social physics : how social networks can make us smarter*.

Petersen, K. *et al.* (2008) 'Systematic Mapping Studies in Software Engineering', *12th International Conference on Evaluation and Assessment in Software Engineering (EASE 2008)*, pp. 71–80.

Phillips, F. (2006) *Social Culture and High-Tech Economic Development*. London: Palgrave Macmillan UK. doi: 10.1057/9780230597242.

Powell, W. W., Koput, K. W. and Smith-Doerr, L. (1996) 'Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Locus of Innovation: Networks of Learning in Biotechnology', *Source: Administrative Science Quarterly*, 41(1), pp. 116–145. doi: 10.2307/2393988.

Programme, T. U. N. D. (2002) *Human Development Report 2001: Making new Technologies work for Human Development*, *The United Nations Development Programme*.

Putnam, R. D. (1993) 'The prosperous community', *The american prospect*, 4(13), pp. 35–42. Available at: http://staskulesh.com/wp-

content/uploads/2012/11/prosperouscommunity.pdf.

Quan-Haase, A. *et al.* (2001) 'Does the Internet Increase, Decrease, or Supplement Social Capital? Social Networks, Participation, and Community Commitment', *Association of Internet Researchers AOIR*.

Ralph H. Kilmann, Mary J. Saxton, R. S. (1987) *Gaining Control of the Corporate Culture*, *Administrative Science Quarterly*.

Ranaei, S. Knutas, A., Salminen, J., Hajikhani, A., 'Cloud-based Patent and Paper Analysis Tool for Comparative Analysis of Research', *CompSysTech '16 Proceedings of the 17th International Conference on Computer Systems and Technologies 2016*, pp. 315–322. doi: 10.1145/2983468.2983490.

Ranaei, S., Suominen, A. and Dedehayir, O. (2017) 'A topic model analysis of science and technology linkages: A case study in pharmaceutical industry', *2017 IEEE Technology & Engineering Management Conference (TEMSCON)*, pp. 49–54. doi: 10.1109/TEMSCON.2017.7998353.

Ravn, T., Nielsen, M. W. and Mejlgaard, N. (2015) *Metrics and indicators of Responsible Research and Innovation*. Available at: https://www.rri-tools.eu/documents/10184/47609/MORRI-D3.2/aa871252-6b2c-42ae-a8d8-a8c442d1d557.

de Reuver, M., Sørensen, C. and Basole, R. C. (2017) 'The digital platform: a research agenda', *Journal of Information Technology*. Palgrave Macmillan UK. doi: 10.1057/s41265-016-0033-3.

Rinkinen, S. (2016) 'Clusters, Innovation Systems and Ecosystems: Studies On Innovation Policy's Concept Evolution and Approaches for Regional Renewal'.

Ritala, P. *et al.* (2013) 'Value creation and capture mechanisms in innovation ecosystems: a comparative case study', *International Journal of Technology Management*, 63(3/4), p. 244. doi: 10.1504/IJTM.2013.056900.

Ritala, P. and Almpanopoulou, A. (2017) 'In defense of "eco" in innovation ecosystem', *Technovation*, 60–61(January), pp. 39–42. doi: 10.1016/j.technovation.2017.01.004.

Robert, L. P., Dennis, A. R. and Ahuja, M. K. (2008) 'Social capital and knowledge integration in digitally enabled teams', *Information Systems Research*, 19(3), pp. 314–334. doi: 10.1287/isre.1080.0177.

Romer, P. (1990) 'Human capital and growth: theory and evidence', *Carnegie-Rochester Conference Series on Public Policy*, (32), pp. 251–286. doi: 10.1016/0167-2231(90)90028-J.

Rose, S. *et al.* (2009) 'Frameworks for Measuring Innovation: Initial Approaches', *Athena Alliance, Washington*, (March), p. 24. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Frameworks+for+Measuring+Innovation+:+Initial+Approaches#0 (Accessed: 22 June 2013).

Rothschild, M. (1990) *Bionomics: Economy As Ecosystem*. Available at: http://www.amazon.com/Bionomics-Economy-Ecosystem-Michael-Rothschild/dp/0805019790.

Rozin, P. and Royzman, E. B. (2001) 'Negativity Bias, Negativity Dominance, and Contagion', *Personality and Social Psychology Review*, 5(4), pp. 296–320. doi: 10.1207/S15327957PSPR0504_2.

Russell, M. G. *et al.* (2011) 'Transforming Innovation Ecosystems through Shared Vision and Network Orchestration', *Triple Helix IX International Conference: Silicon Valley: Global Model or Unique Anomaly?*, pp. 1–21. doi: 10.1017/CBO9781107415324.004.

Russell, M. G. and Smorodinskaya, N. V. (2018) 'Leveraging complexity for ecosystemic innovation', *Technological Forecasting and Social Change*. Elsevier, (January 2016), pp. 1–18. doi: 10.1016/j.techfore.2017.11.024.

Safire, W. (2009) 'CROWDSOURCING', *The new york times*. Available at: http://www.nytimes.com/2009/02/08/magazine/08wwln-safire-t.html?_r=3&ref=magazine&.

Saunders, M. N. K. *et al.* (2012) *Research Methods for Business Students*. Pearson Education. Available at: https://books.google.co.uk/books?id=u4ybBgAAQBAJ.

Schuller, T. (2001) 'The Complementary Roles of Human and Social Capital', *The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being, International Symposium Report edited by the OECD and HRDC.*

Schumpeter, J. (1934) 'The Theory of Economic Development', *Joseph Alois Schumpeter*, pp. 61–116.

Scott, R. and Vincent-Lancrin, S. (2014) *The Global Innovation Index 2014: The Human Factor In Innovation*.

Scott, W. and Davis, G. (2007) 'Organizations and organizing', *Rational, natural and open systems perspectives. New …*, pp. 1–29. doi: 10.4324/9781315663371.

Sebastiani, F. (2001) 'Machine Learning in Automated Text Categorization', *ACM computing surveys (CSUR)*. Available at: https://arxiv.org/abs/cs/0110053v1.

Selden, S. C. and Sowa, J. E. (2003) 'Testing a Multi-Dimensional Model of Organizational Performance: Prospects and Problems'.

Shane, S. (2000) 'Prior knowledge and the discovery of entrepreneurial opportunities', *Management Science*, 47(2), pp. 205–220. doi: 10.1287/orsc.11.4.448.14602.

Shapira, P. *et al.* (2006) 'Knowledge economy measurement: Methods, results and insights from the Malaysian Knowledge Content Study', *Research Policy*, 35(10), pp. 1522–1537. doi: 10.1016/j.respol.2006.09.015.

Sievert, C. and Shirley, K. (2014) 'LDAvis: A method for visualizing and interpreting topics', *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pp. 63–70. Available at: http://www.aclweb.org/anthology/W/W14/W14-3110.

Simon, H. (1996) *The Sciences of the Artificial*, *MIT Press*.

Skolnikoff, E. B. (1993) 'The Elusive Transformation: Science, Technology, and the Evolution of International Politics', *Princeton University Press*.

Small, H. (1988) 'Mapping the dynamics of science and technology', *Scientometrics*, 14(1–2), pp. 165–168. doi: 10.1007/BF02020250.

Smorodinskaya, N., Russell, M. G. and Katukov, D. (2017) 'Innovation Ecosystems vs . Innovation Systems in Terms of Collaboration and Co-creation of Value', pp. 5245–5254.

Stanley, J. and Briscoe, G. (2010) 'The ABC of Digital Business Ecosystems', *Communications Law - Journal of Computer, Media and Telecommunications Law*, 15(1), p. 24. Available at: http://arxiv.org/abs/1005.1899.

Starbird, K. *et al.* (2010) 'Chatter on the red: what hazards threat reveals about the social life of microblogged information', *CSCW '10 Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pp. 241–250. doi: 10.1145/1718918.1718965.

Statista (2017) *Number of social network users worldwide from 2010 to 2021 (in billions)*. Available at: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ (Accessed: 20 January 2018).

Statista (2018) *Most famous social network sites worldwide as of January 2018, ranked by number of active users (in millions)*. Available at: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ (Accessed: 21 March 2018).

Still, K. *et al.* (2014) 'Insights for orchestrating innovation ecosystems: the case of EIT ICT Labs and data-driven network visualisations', *International Journal of Technology Management*. IEEE, 66(2/3), p. 243. doi: 10.1504/IJTM.2014.064606.

Stone, A., Shipp, S. and Leader, P. (2008) 'Measuring Innovation and Intangibles : A Business Perspective', *Innovation*, (April), pp. 532–542. Available at: http://www.athenaalliance.org/pdf/MeasuringInnovationandIntangibles-STPI-BEA.pdf.

Stoneman, P. (2010) *Soft Innovation: Economics, Product Aesthetics, and the Creative Industries*. Oxford University Press. doi: 10.1093/acprof:oso/9780199572489.001.0001.

Sundbo, J. (2006) 'Introduction The Organizational or "Soft" Aspects of Innovation', in Sundbo, J. et al. (eds) *Contemporary Management of Innovation*. Palgrave Macmillan, pp. 125–130. doi: https://doi.org/10.1057/9780230378.

Suominen, A., Seppänen, M. and Dedehayir, O. (2016) 'Innovation Systems and Ecosystems : a Review and Synthesis', in *The XXVII ISPIM Innovation Conference – Blending Tomorrow's Innovation Vintage, Porto, Portugal on 19-22 June 2016.*

Sworder, C. (2017) *The Global Cleantech Innovation Index 2017 Global Cleantech Innovation Programme (GCIP) Country Innovation Profiles*. Available at: https://www.unido.org/sites/default/files/2017-11/GCII_GCIP_report_2017.pdf.

Teece, D. J. (1986) 'Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy', *Research Policy*, 15(6), pp. 285–305. doi: 10.1016/0048-7333(86)90027-2.

Thomas, L. D. W., Sharapov, D. and Autio, E. (2015) 'Linking entrepreneurial and innovation ecosystems: the case of AppCampus', *Entrepreneurial Ecosystems and the Diffusion of Startups*, pp. 35–64. doi: 10.4337/9781784710064.00008.

Tidd, J. and Bessant, J. (2014) 'Strategic Innovation Management', pp. 1–417.

Tidd, J. and Bessant, J. (2007) *Innovation and entrepreneurship*. Available at: https://books.google.com.co/books?hl=es&lr=&id=kKvKh7pla8kC&oi=fnd&pg=PR11&ots=ENhLjHsLXe&sig=4jXp_WT46hV_1cFn8LnAqjzX7Z4&redir_esc=y#v=onepage&q&f=false.

Townsend, A., Pang, A. S.-K. and Weddle, R. (2009) 'Future Knowledge Ecosystems The Next Twenty Years of Technology-Led Economic Development', *Institute for the Future*, pp. 1–40. Available at: http://www.iftf.org/uploads/media/SR-1236 Future Knowledge Ecosystems.pdf.

UNDP (2015) *The global competitiveness report 2015-2016*, *World Economic Forum*. doi: 92-95044-35-5.

Utz, S. and Muscanell, N. (2015) 'Social Media and Social Capital: Introduction to the Special Issue', *Societies*, 5(2), pp. 420–424. doi: 10.3390/soc5020420.

Vaishnavi, V., Kuechler, W. and Petter, S. (2005) 'Design research in information systems', *Association for information systems*. Available at: http://desrist.org/design-research-in-information-systems/.

Valkokari, K. (2015) 'Business , Innovation , and Knowledge Ecosystems: How They Differ and How to Survive and Thrive within Them', *Technology Innovation Management Review*, 5(8), pp. 17–24.

Vieweg, S. *et al.* (2010) 'Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness', *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, p. 1079. doi: 10.1145/1753326.1753486.

Van Aken, J. E. (2005) 'Management research as a design science: Articulating the research products of mode 2 knowledge production in management', *British Journal of Management*, 16(1), pp. 19–36. doi: 10.1111/j.1467-8551.2005.00437.x.

Wardhani, A. R., Acur, N. and Mendibil, K. (2016) 'Human capital, social capital and innovation outcome: A systematic review and research agenda', *2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 355–359. doi: 10.1109/IEEM.2016.7797896.

Wasko and Faraj (2005) 'Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice', *MIS Quarterly*, 29(1), p. 35. doi: 10.2307/25148667.

Weikum, G. (2002) 'Foundations of statistical natural language processing', *ACM SIGMOD Record*, p. 37. doi: 10.1145/601858.601867.

White, H. D. and McCain, K. W. (1998) 'Visualizing a discipline: An author co-citation analysis of information science, 1972-1995', *Journal of the American Society for Information Science*, 49(4), pp. 327–355. doi: 10.1002/(SICI)1097-4571(19980401)49:4<327::AID-ASI4>3.0.CO;2-W.

Wiebe, J. *et al.* (2003) 'Recognizing and Organizing Opinions Expressed in the World Press.', *In Working Notes - New Directions in Question Answering (AAAI Spring Symposium Series*, pp. 12–19. Available at: http://www.aaai.org/Papers/Symposia/Spring/2003/SS-03-07/SS03-07-003.pdf.

Wilkins, J., Van Wegen, B. and De Hoog, R. (1997) 'Understanding and valuing knowledge assets: Overview and method', *Expert Systems with Applications*, 13(1), pp. 55–72. doi: 10.1016/S0957-4174(97)00022-5.

Williams, D. (2006) 'On and Off the 'Net: Scales for Social Capital in an Online Era', *Journal of Computer-Mediated Communication*, 11(2), pp. 593–628. doi: 10.1111/j.1083-6101.2006.00029.x.

Wolfe, R. A. (1994) 'Organizational innovation: Review, critique and suggested research directions', *Journal of Management Studies*, 31(3), p. 405.

World Economic Forum (2017) *Mapping Global Transformations*. Available at: https://www.weforum.org/about/transformation-maps.

Xu, G. *et al.* (2017) 'Exploring innovation ecosystems across science, technology, and business: A case of 3D printing in China', *Technological Forecasting and Social Change*. doi: 10.1016/j.techfore.2017.06.030.

Yi, J. *et al.* (no date) 'Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques', *Third IEEE International Conference on Data Mining*, pp. 427–434. doi: 10.1109/ICDM.2003.1250949.

Yin, R. K. (2013) 'Case Study Research: Design and Methods', p. 303. doi: 10.1097/FCH.0b013e31822dda9e.

Zhao, F. (2005) 'Exploring the synergy between entrepreneurship and innovation', *International Journal of Entrepreneurial Behavior & Research*, 11(1), pp. 25–41. doi: 10.1108/13552550510580825.

PART II: INDIVIDUAL PUBLICATIONS

Publication I

Hajikhani, A

Emergence and Dissemination of Ecosystem Concept in Innovation Studies: A Systematic Literature Review Study

Reprinted with permission
In Proceedings of the 50th Hawaii International Conference on System Sciences
ISBN: 978-0-9981331-0-2
DOI: 10.24251/HICSS.2017.634

Emergence and dissemination of ecosystem concept in innovation studies: A systematic literature review study

Arash Hajikhani
Lappeenranta University of Technology
arash.hajikhani@lut.fi

Abstract

The abundant academic literature on ecosystem approach has attracted the attention of many researchers in various fields (such as business and management, innovation and entrepreneurship). The debate is questioning whether ecosystem as a concept is going towards a matured path. This article, contributes to this debate by applying bibliometric analysis approach on two similar comparable concepts of "system" and "ecosystem" within the boundaries of Business and Management (B&M) discipline. The co-citation analysis shows that the "system" concept is structured on a stable body of references and disseminated into various fields and subdomains. While the "ecosystem" concept is at early development phase, and it is expected to be identified as a distinguished field which provides explicit added value to innovation related research community. This paper provides the network structure of the collaborating authors, compares the disseminating pattern of the concepts and performs an advance topical analysis of respected literatures.

1. Introduction

The term ecosystem has been used with a substantial interest by scholars in field of innovation management [1]. The term has been borrowed by variety of scientific domains and indeed has been co-opted in the press in the past few years to describe various phenomena [13,15].

Ecosystem is a derived term from system. The difference between ecosystem and system is that an ecosystem is a community of living organisms in conjunction with the nonliving components of their environment (things like air, water and mineral soil), interacting as a system [4]. In simple words, ecosystem is a system formed by an ecological community and its environment that functions as a unit, while system is a collection of organized things; as in a solar system [4].

The ecosystem concept has roots in ecology, where it claims that it consists of the biological community that occurs in some local, and the physical and chemical factors that make up its non-living or abiotic environment. Ecology is branch which studies living organisms and their interaction with the environment they inhabit [14].

The term has been utilized in fields of business and economics by Rothschild in 1990 [17]. In Rothschild's book "Bionomics", he is promoting the understanding of biology in direction of understanding our economic future. He points to ecological dimensions of economy and elaborates interesting parallel and analogies between business and biology. Bionomics is defined as the branch of ecology that examines the economic relations between organisms (organizations) and their environment [17]. The bionomic perspective illuminates the interaction of forces that maintain stability while spawning changes. Later on, Moore in 1993 [11] took the ground and introduced the term "Business ecosystem" by which he emphasized the essentiality of competition among ecosystem components. The author further stressed the dynamics that regenerate the interactions between organisms and the environment.

Previous works have defined and distinguished the concept of ecosystem and gave it a framework [6,8,10]. While, the "ecosystem" phrase itself was not very successful in embedding itself in new literature, it has also been criticized for lack of clarity[13]. The emergence of the concept "ecosystem" in business and management studies, has attracted scholars attention toward tracking this growth and exploring the dissemination of the concept to other fields. The major challenge of any noble and emerging concept is to define itself and disseminate to other disciplines and domains of study. The contribution of this paper is exploring the operationalization ,impact

URI: http://hdl.handle.net/10125/41796
ISBN: 978-0-9981331-0-2

assessment of the ecosystem concept and its further dissemination to other disciplines.

One way to analyze the creation of knowledge and its diffusion in an emerging field is to use references co-citation analysis [9,20]. References co-citation analysis is a useful approach when it comes to exploring the knowledge structure of a research domain [21]. This analytical technique also serves to discover knowledge diffusion and influence among a research community. It sheds light on the networks of references, the social construction of a field, and on its intellectual advances. Yet, co-citation analysis does not directly provide insights on future trends. This research has leveraged the advancements in bibliometrics data analytics (refers to statistical analysis of written publications) for exploring the evolutionary path of the concept of ecosystem in B&M studies.

To contribute to the debate on weather or not the ecosystem concept has been able to establish itself in innovation studies within the boundaries of B&M discipline, this paper take comparison into the context and decided to take into consideration the "system" concept for benchmarking purposes. The initial investigation showed that both of the concepts (system and ecosystem) have been widely adopted in innovation studies literature, therefore it would be sensible practically and contextually to limit the scope to the innovation studies domain. It will be investigated whether ecosystem has the characteristics of a concept or approach through a bibliometric analysis of the literature within the boundaries of B&M discipline. On the other spectrum, the concept of system will be investigated to see its usage and adoption overtime since it is a well-established concept and close to the meaning of ecosystem. The comparison of system and ecosystem concept in B&M discipline will provide a fair comparison ground to observe the dissemination trend of the concepts.

To this end, the literature which adopted the concepts of system and ecosystem within boundaries of B&M discipline (characteristics such as: major publication venues, main used keywords, influential papers) will be identified. Second, the structures of the literature and the most inflectional scientific articles as the core literature for each of the concepts will be explored. An analysis will be conducted to evaluate the network structure and density of the core literature for each concept. Later on, the analysis will be escalated by encountering the co-citations of the papers which have often cited the core literature and by analyzing their respected domains, the areas which the concepts have been disseminated will be discovered.

The paper is organized as follows; second section presents the methodological approach. Then the bibliometric analysis will be outlined based on the procedure which is described in the methodology section. Thereafter, the findings, discussion and conclusions are presented, the final section.

2. Methodology and data

In this section, the method used to identify and analyze the bibliometric data will be presented (consisting of: title, abstract, year of publication, authors, publication venue, keywords, list of references). The searching queries and data collection process will be explained in detail as well.

2.1. Data collection

The concept of ecosystem and system which has been adopted in the B&M discipline will be analyzed. Thereby, the focus is on research articles that addressed the terms or variation of the terms in their titles. For that purpose, Web of Science (previously known as (ISI) Web of Knowledge) has been used as a searching database that includes 90 million documents indexed and is considered to be one of the most important databases for scientific bibliometric data. The Web of Science (WoS) core collection will be incorporated to enrich the coverage to all type of indexed documents.

2.2 Methodology

Bibliometric data analysis was conducted as a means to provide quantitative analysis of academic literature [12]. Bibliometrics is known as statistical analysis of written publications and citation analysis. The bibliometric method is based on constructing the citation graph, a network or graph representation of the citations between documents. Many research fields use bibliometric methods to explore the impact of their field, set of researchers, or a particular paper [16].

The bibliometric data analysis was facilitated by the help of toolkit for Network Analysis Interface for Literature Studies "NAILS" that has been developed and published via a conference paper by 2015 [7]. The motivation for using NAILS was to promote its usability and availability as the only open-source cloud based toolkit for bibliometric analysis. Despite of expensive commercialize, closed system tools which are required to be setup and need expert knowledge in data preparation and processing, NAILS proposed an open, extensible tool with even

more automated workflows which will make this bibliographic analysis available to a wider part of the community of researchers. The literature analysis tool "NAILS", which uses a series of custom statistical and network analysis functions, offers its users an overview of literature datasets retrieved from WoS. (access from: http://nailsproject.net).

The overall process which is conducted for the systematic literature review for the concepts of system and ecosystem is illustrated in Figure 1.

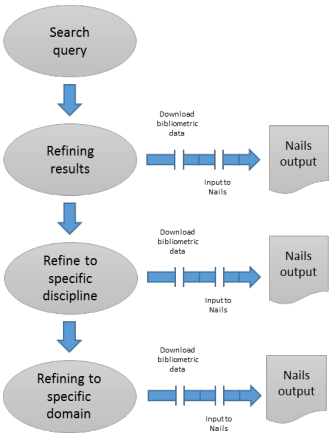


Figure 1. Steps for conducting bibliometric analysis with Nails

The steps were accordingly:

Step 1: Scientific information retrieved from targeted database (Thomson Reuters Web of Science Core Collection), which was utilized for building the search query with initial keywords. The query was built by using the keywords and boolean operators (e.g. *innovate* AND (ecosystem* OR eco-system*)*) and then executed.

Step 2: After the query search in the WoS core collection, initial refining (including and excluding of the records) have been done for the result records.

Step 2.1: The refined dataset including the citation data was downloaded in tab-delimited format from WoS manually.

Step 2.2: The downloaded bibliometric data was bundled by a compression tool (option for compression is available in Mac and Windows) and was uploaded onto the Nails (http://hammer.nailsproject.net) online interface.

Once the analysis has been initiated on NAILS, new metrics were calculated such as PageRank (It counts the number and quality of links to a paper to determine a rough estimate of its importance) and In-Degree (Provides the number of citations coming into a paper in a directed graph) on the citation data of the records. As part of the new metrics, a tailor-made report was generated that provides an abstract/keyword analysis, productive authors/journals and gives recommendations for including top publications based on the citation data. In addition, required data files were also generated in order to graph the network of the records visually.

The Steps 3 and 4 happened as the goal is to dig into a particular domain of study in the concerned discipline. The following sub steps for step 3 and 4 are preceded in the same way as described for the step 2. The bibliometric data from scientific publications were further leveraged for a more extensive and accurate literature analysis. In order to investigate the dissemination of the concerned research domains, the core literature has been detected so to see how often and in what rate the core literature has been cited.

Detecting the "core literatures" is one of the effective ways for distinguishing impactful papers in a concerned domain of study and its relevant literatures [3]. Core literature or documents are advantageous to identify further relevant documents by following the formers' strong and medium-strong links. The notion of core literature was first presented by Small [19] in connection with co-citation analysis. The concept has been escalated by Glänzel and Czerwon [5] on the basis of bibliographic coupling to identify literatures which form important nodes in the network of scholarly communication. In general, the focus on bibliometric analysis is on the citation networks of individual publications. Cooper et al. [3] showed that citation connections could express the relevance to the topic of discussions. Therefore, if a set of records is more highly cited by other publications in a certain domain field, then these records have a greater possibility of belonging to the same domain field.

The interpretation of "core literature" in this paper, represents the most related and impactful papers in the concerned domain of study. Meanwhile, they might not necessarily be interlinked as the concerned discipline might be an emerging one or the topic is highly multidisciplinary in nature. The core literature was then utilized to identify documents which have often cited the core literature. Figure 2 is a good representation of the definition of "core literature" illustrated in this research

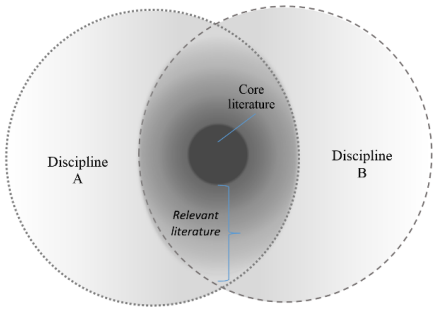


Figure 2. Visualization of the core literature

The process of defining the core literature is manual and the main target is to define the relevancy to the core literature as proxy for filtering the relevant articles. Figure 3 is an illustration of how the practice has been utilized schematically.

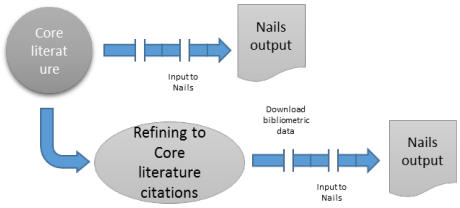


Figure 3. Core literature analysis

After distinguishing the core literature, the records were downloaded and transferred to NAILS for an initial overview (compare with the sub steps 2.1 and 2.2 which were described previously).

The process for collecting the papers which have cited the core literature was manually extensive to conduct, but it is necessary in order to see the dissemination pattern of the core literature. The process includes collecting the full bibliometric data from these references. By retaining the citation's bibliometric data that cited the core literature, it would be clear which papers have cited the most of the core literature and in general to what fields they have been disseminated.

Extracted citations get analyzed within the core literature in NAILS, new indexes were calculated in the NAILS report, which shows an indication of the relevancy of the records in regards to the core literature. The fact that the citations to the core literature depends on availability of them relative to the publication time has been encountered. Therefore, the ratio has been developed that highlights the relevancy of the records to the core literature based on the number of citations which has been made in the papers (Formula 1). "Times cited per year" is another indication which illustrates the quality of the paper based on the average citation which it gets each year.

$$Relevance\ index = \frac{\text{Number of Citation to Core Literature}}{\text{Number of Available Core Literature}} \times 100$$

Formula 1. Relevance index

Following section, the applied procedure will be utilized to understand how the system and ecosystem concepts have been adopted in B&M discipline. A detailed comparison will be constructed for the usage of the two concepts of system and ecosystem within the context of innovation. The detection of the core literature for the two concepts is necessary to understand the dissemination of the concepts into other study domains.

3. System concept in business and management discipline

Here, the practice is to show the usage of system approach in B&M discipline and how innovation studies adopted the concept. The search query was built using the keywords and boolean operators and wildcard like "*" (The use of asterisk (*) as a truncation symbol allowed the databases to look for different endings of the word). The search executed in the title. Usage of system and its variation (i.e. system/s, systematic/s, systemic/s) in WoS core collection for English language has been looked at which the initial results ended up with 1,844,467 records.

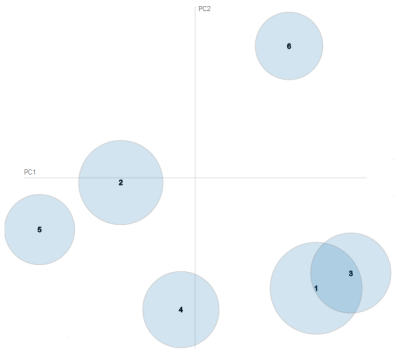
Search words in Title	system*
Time span	August 2016 All years WoS core collection: **Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.
Search refining criteria	English language

According to WoS subject categories, more than 50% of the results are dominated by categories such as electrical engineering, automation, computer science in artificial intelligence, information systems and telecommunications. Refining the initial results into B&M discipline (the domain isolation was

facilitated by using WoS subject categories) led to 70,083 records.

Due to the big size of the results, only top 3,000 cited publication has been analyzed with the NAILS toolkit in order to have a better overview of the usage of the concept of system in B&M studies. The keyword analysis indicated; decision support and information systems, supply chain management and simulation as the top popular keywords and user satisfaction, knowledge management and innovation as highly cited keywords (The full report can be found from this link online).

Topic modeling technique has been applied to analyze the abstracts contents. The technique is a type of statistical model for discovering the abstract "topics" that occur in a collection of documents in order to explore hidden semantic structures in a text body [2]. More precisely, the visualized application of the "Latent Dirichlet allocation" introduced by Sievert and colleagues [18] was utilized in order to perform the topic generation of the analyzed abstracts. Figure 4 is an illustration of the popular distant topics/themes related to the system concept used in B&M discipline. (The interactive visualization for the topical abstract analysis is available from this link online).



Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
linear	decision	market	research	production	research
feedback	fuzzy	health	organizational	optimal	innovation
stability	approach	economic	user	inventory	knowledge
nonlinear	selection	financial	information	problem	management
control	method	costs	success	cost	development
state	dss	firms	support	demand	systems
time	proposed	environmental	organizations	scheduling	business
problem	process	risk	findings	policy	technology
systems	making	countries	perceived	manufacturing	organizational
matrix	based	cost	users	maintenance	social

Figure 4. Abstracts topical clustering for system concept in B&M

The domain boundaries have been narrowed to innovation studies by adding another criterion "innovate*" into the searching query. The results became 2,140 records that had applied the system approach and were within the boundaries of B&M and contain the context of innovation.

The top popular keywords were: Innovation, regional and national innovation system. Whereas, triple helix (university-industry-government relations) was among the top highly cited keywords after Innovation and innovation system. The top 25 publications have high number of in-degree which shows the interconnectivity of the concept of system in the innovation studies literatures (The full report can be found from this link online).

The process for detecting the core literature was initiated. The 2,140 records were organized and selected according to the PageRank, In-Degree and number of accumulative citation. The title and abstract of the top ranked paper have been read in order to make sure they are within the subject. 67 core literature studies dealing with system concept in business and management with the context of innovation were selected (The NAILS analysis report for core literature from this link online).

The analysis is continued by collecting the bibliometric data of papers which have cited the core literature for the purpose of 1) defining the relevant papers who adopted the concept; 2): understanding the dissemination of the concept into other disciplines. Finally 7,225 full bibliometric data and all the citation of the core literature were extracted (The NAILS analysis online report for the citations to the core from this link online).

Processing the extracted citations analyzed within the core literature in NAILS, Relevance index was calculated in NAILS report, which shows an indication of the relevancy of the records regarding to the core literature. The minimum criteria have been considered for relevancy to core literature to be at least 2 times reference to the core literature. With that criteria, the result ended up to be 1,593 records which have cited the minimum of 2 of the articles in the core literature. To illustrate the dissemination impact, the number of 19,102 citations has been carried out by the 1,593 paper with minimum 2 reference to the core.

Figure 5 shows the distribution of the generated citations based on number of times the papers (1,593) have been citing the core literature. For example, the papers which have 2 of the core literature in their references (a proxy for relevancy), have generated 10,756 citations in total.

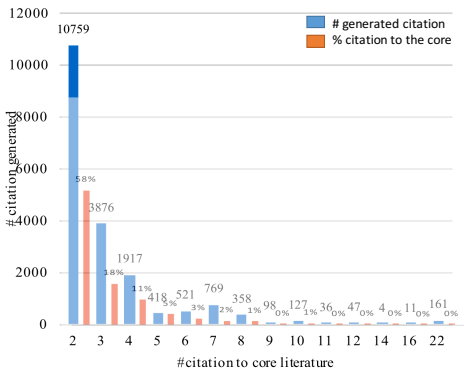


Figure 5 number of citations generated by papers citing the core literature

In Figure 6 an attention has been paid at the subject categories which the papers citing the core literature have managed to penetrate.

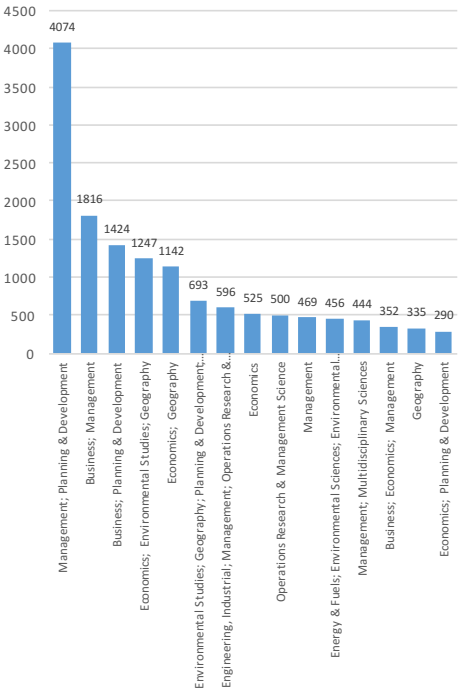


Figure 6. Distribution of the subject category of the core literature citations

As it can be seen from Figure 6, business and economics with various orientations were generating the highest amount of citation and impact, in relative terms. It is also visible that the dissemination of the core concept had an effect in Environmental sciences, Geography and industrial engineering.

4. Ecosystem concept in business and management discipline

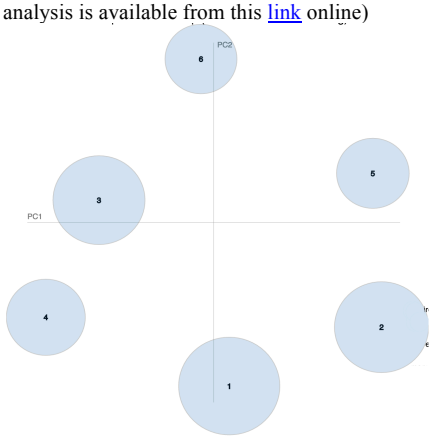
Looking up the term "ecosystem" and its variations in WoS core collection database, retrieved 38,940 results. A descriptive analysis on the results identified, that the majority use of the term is in ecology, biology, oceanography, forestry and environmental discipline while business and management affiliated materials represents only 6.3% of the documents. There is much conceptual ambiguity surrounding ecosystems as it had been discussed in introduction. Ecosystems are a metaphor, taken from biology, which is often ill-defined. Ecosystem is highly fashionable label therefore its important to notice the underling phenomenon which might be very similar. Other terms have been used extensively to capture ecosystem concept which is needed to be taken into account. Therefore, in order to not to focus on label "ecosystem", the term ecosystem should be expanded. In order to achieve that, terms which associate with the concept of ecosystem are required to be collected. In this regard, a descriptive analysis should be run on the publications which have the ecosystem term in titles in B&M discipline to see which term associate with ecosystem. The search in WoS for looking up the term ecosystem (i.e. eco-system*, eco system* and ecosystem*) in WoS core collection database retrieved 643 records. By running a NAILS analysis on the retrieved bibliometric data, it has been noticed that among the popular keywords, terms such as platform, value network, innovation network, quadruple helix and mode 3 innovation ecosystem exists. The keywords were incorporated in the WoS search query.

Search words in Title	(ecosystem*) OR ("eco system*") OR (eco-system*)OR (platform*) OR ("value network*") OR ("quadruple helix") OR ("innovation network*") OR ("mode 3 innovation ecosystem*")
Development day	August 2016
Time span	All years
Databases	WoS core collection: **Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH,

	BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.
Search refining criteria	English language

The search resulted to 3,260 records in B&M discipline. The keyword analysis with NAILS indicated popular keywords such as: ecosystem services, innovation, platform, innovation networks and most cited keywords as: ecosystem services and valuation (The NAILS analysis report from this link online).

Figure 7 illustrates the abstract topical cluster analysis for the publication related to ecosystem concept in B&M discipline boundaries. (The interactive visualization for the topical abstract analysis is available from this link online)



Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
innovation	product	ecosystem	platform	platform	knowledge
network	platform	services	system	platforms	business
knowledge	design	ecological	teaching	market	ecosystem
networks	manufacturing	water	information	online	research
regional	platforms	environmental	service	consumers	ecosystems
industrial	system	land	learning	media	social
enterprises	simulation	economic	logistics	users	service
industry	approach	ecosystems	students	competition	creation
enterprise	proposed	conservation	commerce	effects	understanding
development	time	values	web	markets	processes

Figure 7. Abstracts topical clustering for ecosystem concept in B&M

Analyzing the abstracts of the papers with topic modeling (LDA), illustrated that the popular distant topics/themes which ecosystem have been used in B&M discipline. Ecosystem related literature is apparent in different topics (1,2) while platform related literature is apparent in topics (3,4,5,6). Innovation topic is shared with both ecosystem and platform only in topic 1 (which has a healthcare theme).

The study domain boundaries narrowed to innovation studies by adding another criterion "innovat*" into the searching query. The results became 988 which represents the records that have applied ecosystem approach and are within the boundaries of B&M and contain the context of innovation (The full report can be found from this link online).

The popular keywords were: innovation ecosystem and its combination obviously, open innovation, entrepreneurship, learning processes, collaboration and knowledge management. Whereas the highly cited keywords were: business ecosystem, vertical integration, technological change ecosystem services, social and traditional media, online ecosystems and marketing metrics.

The top 50 publications have relatively much lower In-Degree ratio within the publication that represents the low interconnectivity of literature. For the purpose of detecting the core literature, the 988 records were organized and selected according to the PageRank, In-Degree and number of accumulative citation. The title and abstract of the top ranked paper has been read in order to make sure they are within the subject. 42 core papers dealt with the ecosystem concept in innovation studies within the boundaries of B&M study (The full report can be found from this link online).

The analysis continues by collecting the papers which have cited the core literature and accordingly their bibliometric data. The purpose for this was to, 1): defining the relevant papers who have adopted the concept, and 2): understanding the dissemination of the concept into other subject categories. 5,335 full bibliometric data were extracted (The full report can be found from this link online). Considering the minimum of 2 references to the core literature, 286 papers have meet the criteria. The distribution of the relevant papers regarding the number of times they have cited the core literature is illustrated in Figure 8.

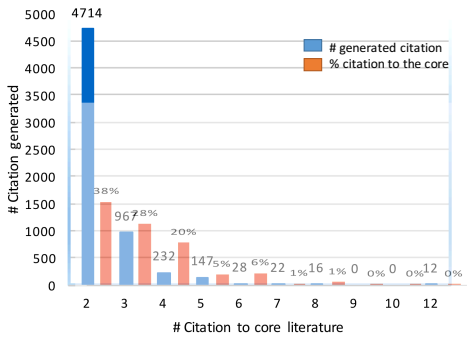


Figure 8. number of citations generated by papers citing the core literature

In Figure 9, subject categories have been observed so to see the penetration of the core literature.

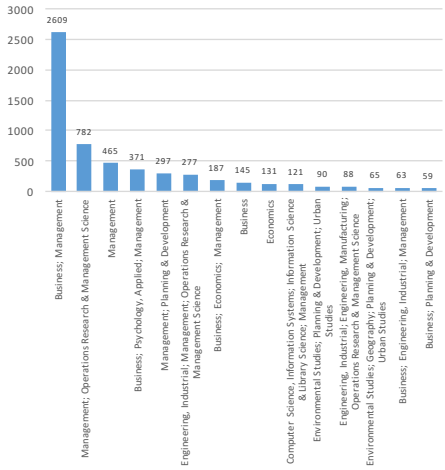


Figure 9. Distribution of the subject category of the core literature citations

The disseminated subject categories which received citation and generate impact are business, management and economy and operation research. The notion of ecosystem within business and management discipline had defused to areas such psychology, computer and information science, environmental, urban and life sciences as well as industrial and engineering.

5. Findings

The citation network structure of the two concepts of system and ecosystem in innovation studies within the boundaries of business and management discipline has been reviewed. In the initial phase, the required bibliometric data have been searched and retrieved then based on the defined process the cloud based toolkit for bibliometric analysis "NAILS" had been leveraged to perform the analysis. The investigation was meant to compare the performance and network structure of system and ecosystem concept in innovation studies.

The magnitude of the usage of system concept was 21 times bigger comparing to ecosystem concept in all publication outlets in business and management discipline. Getting to innovation studies, an assessment of the adoption of both concepts revealed that the system concept was only 2 times bigger than the ecosystem concept which is an indication of a relatively bigger contribution of the latter concept into the innovation study domain.

The topical analysis of the abstracts revealed that, papers applying the ecosystem concept are covering more distinct topics than system concept related papers. Comparing the Figures 4 and 7, the six topics clustered papers applying the ecosystem concept have relatively far distance from each other which implies higher diversity in the literature, whereas the topical cluster for the system concept papers have closest distance by each other.

The comparison of dissemination of the two concepts are very informative as it clearly can be seen that the system concept has been disseminated in planning developing and environmental studies in higher extend comparing to ecosystem concept. On the other hand, ecosystem concept was superior in disseminating to areas such as psychology, computer and information sciences and urban studies. The dissemination pattern for the system concept in innovation studies are mainly generated by publications with 2 referred core literature while for ecosystem concept 3 and 4 referred papers has higher portion. The overall dissemination of the system concept in citation terms translated to be 750 papers from core literature which generated over 19,102 citations. The dissemination of the concept of ecosystem to the literature is counted as 286 paper which managed to generate 6,138 citations.

Moreover, an analysis performed on author's network for the both concepts which is shown in the Figures 10 and 11. The network is consisting of nodes which represent authors and edges which represents collaboration or coauthor ship.

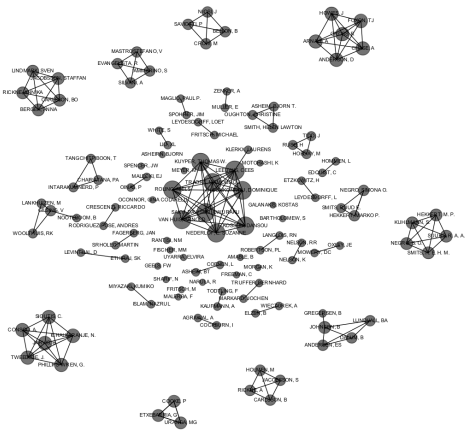


Figure 10. System concept in innovation context

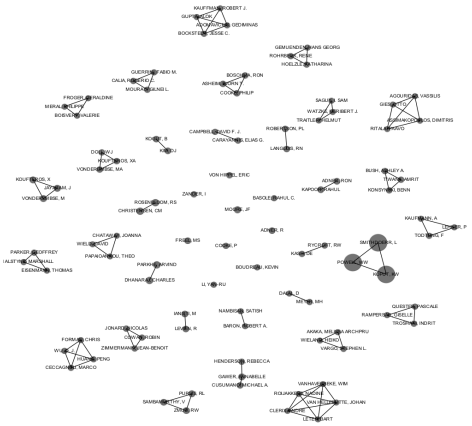


Figure 11. Ecosystem concept in innovation context

Comparing the network structures, it can be observed that less relative collaboration is happening among authors researching the ecosystem concept in the core literature than those focusing on the system concept in the core literature. Translating these patterns to numbers, the average degree (The mean amount of connections per node on the graph) is 2.484 for Figure 10 whereas for Figure 11 equals to 1.835. The other important metric is graph density which is a ratio of the number of edges per node to the number of possible edges. The graph density ratio for Figure 10 is 0.02 while for Figure 11 is lower at number 0.015. The reported ratio indicates that although there exist highly cited individual nodes in the ecosystem concept core literatures (high In-Degree value of nodes), the authors were not able to find each other and collaborate. The result of low level of collaboration for the ecosystem concept leaders is eminent while looking at dissemination pattern of the articles citing the core literature. The diversified and less cited escalated articles are a result of the missing depth and conceptual ground work from the ecosystem concept core literature. It is very dominant that lack of unity in the knowledge base of ecosystem concept in B&M caused by the variety usage of keywords (e.g. ecosystem, open innovation, entrepreneurship, collaboration, platforms, and networks) which weakens the momentum for the diffusion of the concept.

A further implication from NAILS analysis report of the bibliometric data is that the system concept has focused on communities and venues for publication, while the ecosystem showed a dispersed behavior on appearing in publication venues.

6. Discussion and conclusion

This study has contributed to the field of innovation management literature in several ways. First, it introduces a structured approach for analyzing bibliometric data with an orientation of tracking a concept dissemination. This approach offers a new perspective for understanding how a concept or theory has been disseminated and what the patterns of the author's network are. Second, this study offered a methodological approach into an ongoing debate regarding the system vs ecosystem concepts in an innovation studies context, thereby the attempt is to look at the structure of bibliometric data and citation network.

The analysis in this paper presents a deeper understanding of the usage of the concepts of system and ecosystem in business and management as a discipline, by interpreting their bibliometric characteristics, and determining the current maturity of the fields based on their dissemination orientation.

Comparing to a system as a concept, the initiation of the concept of ecosystem in B&M studies was carried out with a lack of consistency and interconnectivity of authors. One explanation for the fragmentation of author's collaboration might be caused by the new terms usage that eventually removes the connection with older same concept publications. This fact of moving to a new word usage influences the citation in which it departures and loose of origins. This paper suggests that it is important to develop a commonly understood ecosystem vocabulary that allows a comparison among studies. Furthermore, a shortage of micro level case studies to illustrate the usage of the designed frameworks is suggested. A comparison of such case studies looking into innovation system vs innovation ecosystem would help to differentiate the concepts clearly. Further research on innovation and ecosystem would ideally investigate in more detail, what ecosystem concept approach is needed in different situations. The concept evolutionary path should be guided in order to identify theoretical approaches, such as, principles, indices, models, frameworks, and tools. These approaches would form the necessary foundation for future empirical research and theory development and validation.

It is concluded that ecosystem as a newly emerging concept in innovation studies is maturing by the number of publications, but it is still not attached to an epistemological orientation. Considering the maturity of the usage of system in innovation studies within the boundaries of B&M, this study suggests the use of a unified definition and

metrics and calls for a collaboration with the authors within the already established community. Concurrently, while the concept ecosystem might add complexity to the current understanding of the system concept that currently dominates B&M, the former adds a new perspective or at least pinpoint the aspects which were underestimated previously. it is hoped that this review invites researchers to initiate more rigor research helping to expand our understanding of the concept of innovation ecosystem.

7. References

[1] Erkko Autio and D.W.Thomas Llewellyn. 2014. Innovation ecosystems: Implications for innovation management. *The Oxford handbook of innovation management*, January: 204–228. http://doi.org/10.1093/oxfordhb/9780199694945.013.012

[2] David M. Blei. 2012. Probabilistic topic models. *Communications of the ACM* 55, 4: 77. http://doi.org/10.1145/2133806.2133826

[3] Harris Cooper, Larry Hedges, and Jeffrey Valentine. 2009. *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.

[4] MW Dictionary. 2015. Merriam-webster online. Retrieved June 5, 2016 from http://www.merriam-webster.com/dictionary/learning\nhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Merriam-Webster+Online#1

[5] Wolfgang Glänzel. 2015. Bibliometrics-aided retrieval: where information retrieval meets scientometrics. *Scientometrics* 102, 3: 2215–2222. http://doi.org/10.1007/s11192-014-1480-7

[6] DJ Jackson. 2012. What is an Innovation Ecosystem. *Arlington, VA*: National Science Foundation http://doi.org/10.1017/CBO9781107415324.004

[7] Antti Knutas, Arash Hajikhani, Juho Salminen, Jouni Ikonen, and Jari Porras. 2015. Cloud-based Bibliometric Analysis Service for Systematic Mapping Studies. *Proceedings of the 16th International Conference on Computer Systems and Technologies*, 184–191. http://doi.org/10.1145/2812428.2812442

[8] Thierry Isckia and Denis Lescop. 2013. Platform- based ecosystems: Leveraging Network- Centric Innovation. November.

[9] Katherine W McCain. 1990. Mapping Authors in Intellecual Space: A Technical Overview. *Journal of the American Society for Information Science* 41, 6: 433–443.

[10] Birol Mercan and Deniz Götkas. 2011. Components of Innovation Ecosystems. *International Research Journal of Finance and Economics* 76, 76: 102–112. http://doi.org/1450-2887

[11] J. F. Moore. 1993. Predators and prey: a new ecology of competition. *Harvard Business Review* 71, 3: 75–86. http://doi.org/Article

[12] Jeppe Nicolaisen. 2010. Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics. *Journal of the American Society for Information Science and Technology* 61, 1: 205–207.

[13] Deog-Seong Oh, Fred Phillips, Sehee Park, and Eunghyun Lee. 2016. Innovation ecosystems: A critical examination. *Technovation*. http://doi.org/10.1016/j.technovation.2016.02.004

[14] Theo Papaioannou, David V. Wield, and Joanna C. Chataway. 2009. *Knowledge ecologies and ecosystems? An empirically grounded reflection on recent developments in innovation systems theory*. http://doi.org/10.1068/c0832

[15] Vaida Pilinkienė and Povilas Mačiulis. 2014. Comparison of Different Ecosystem Analogies: The Main Economic Determinants and Levels of Impact. *Procedia - Social and Behavioral Sciences* 156, April: 365–370. http://doi.org/10.1016/j.sbspro.2014.11.204

[16] Dr Alan Pilkington. 2009. Bibliometrics. *Royal Holloway*. Retrieved from http://personal.rhul.ac.uk/uhtm/001/BibliometricsIndex.html

[17] Michael Rothschild. 1990. *Bionomics: Economy As Ecosystem*. Retrieved from http://www.amazon.com/Bionomics-Economy-Ecosystem-Michael-Rothschild/dp/0805019790

[18] Carson Sievert and Kenneth Shirley. 2014. LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*: 63–70. Retrieved from http://www.aclweb.org/anthology/W/W14/W14-3110

[19] H Small. 1973. Cocitation in Scientific Literature - New Measure of Relationship Between 2 Documents. *Journal of the American Society for Information Science* 24, 4: 265–269. http://doi.org/10.1002/asi.4630240406

[20] H. Small. 1988. Mapping the dynamics of science and technology. *Scientometrics* 14, 1–2: 165–168. http://doi.org/10.1007/BF02020250

[21] Howard D. White and Katherine W. McCain. 1998. Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science* 49, 4: 327–355. http://doi.org/10.1002/(SICI)1097-4571(19980401)49:4<327::AID-ASI4>3.0.CO;2-4

Publication II

Hajikhani, A., Suominen, A.

General System Theory Attributes in Innovation Ecosystem Research Landscape: A Bibliometric and Content Analysis of the Literature

Submitted to Journal of Technological Forecasting and Social Change

General System Theory Attributes in Innovation Ecosystem Research Landscape: A Bibliometric and Content Analysis of the Literature

Abstract

Research on the "innovation ecosystem" has been active for the last 10 years and continues to garner interest from both the scholarly and practitioner communities. The broad-based interest has caused the growth in literature, but simultaneously the cumulative knowledge has become fragmented. This study highlights through the application of General Systems Theory (GST), how we can identify and examine five major system attributes in the emerging ecosystem approach in the literature. Systematic literature review and bibliometric data analysis identifies the core literature. Applying GST reveals the lack of clarity and elaboration on dynamism and goal attributes. The overall nature of the research is exploratory and considers as the first attempt to identify the intellectual structure of GST in the emerging ecosystem approach within innovation studies. Scientific paper's full text analysis utilized in parallel with bibliometric methods to provide an accurate insight of the concept development. Finally, the review and analysis of the literature closes by discussing future research avenues in innovation ecosystems, suggesting implications for researchers and practitioners.

Keywords: General system theory; Innovation ecosystem; Bibliometric analysis; Content analysis; Systematic literature review

1. Introduction

The availability of scientific papers in full text and in machine-readable formats has become more and more widespread thanks to platforms such as Web of Science, Scopus and Google Scholar. In particular, the field of scientometrics is concerned with "the quantitative aspects of the generation, propagation, and utilization of scientific information" (Braun et al. 1987). Bibliometrics traditionally relies on the analysis of metadata of scientific papers and citation analysis is considered as an important scientometric technique to measure and evaluate a publication impact (Vinkler 2010). Due to this structured attempt for indexing the scientific papers, the rise of open access toolkits such as VOSviewer and NAILS have made bibliometric and citation analytics free, accessible and easy to perform. Scholars in many knowledge domains rely heavily on scientometric techniques to search for and retrieve records and publications pertinent to their research interests. Scintometrics is crucial to reduce complex research areas to more relevant and small sets of preselected document set. Yet the problem is complicated when hundreds of documents are meeting the criteria of a search. In addition, unstructured nature of most documents which, unlike databases will increase the difficulty to systematically approach relevant records. Bibliometric and citation analysis methods are proved valuable tools to monitor and chart scientific processes, yet several criticism of citation based measures has been pointed out (Bornmann and Marx 2014; Eysenbach 2011; Gorraiz et al. 2014; Jonkers et al. 2014; Waltman et al. 2013). Up to now full-text mining efforts on scientific papers are rarely utilized to provide data for enhancing bibliometric analyses insights. While bibliometrics traditionally relies on the analysis of metadata of scientific papers, in this research we explore the ways in which bibliometric methods can benefit from full-text analysis of scientific papers. While moving to full-text creates a much richer dataset, but simultaneously calls for new ways of analysis. To structure the complex set created by the full-text, we apply GST to uncover attributes embedded to the full-texts. We focus on a case study on the emerging concept of "innovation ecosystems" in order to study the semantic dimensions of the full text published records in addition to the bibliometric data insights.

Research on "innovation ecosystems" begun roughly ten years ago and continues to garner burgeoning interest from both the scholarly and practitioner communities. Ecosystem as a concept is used as biological metaphor for innovation in an attempt to explain the interaction dynamics of the relationships formed between various components of the innovation system whose functional goal is to maximize the efficiency of the system as a whole (J. F. Moore 1993; Rothschild 1990). Innovation ecosystem has been introduced from business and management (B&M) perspectives such as strategy, innovation management, collaborative networks and clusters, which are defined as a network of collaborative components complimenting and competing each other for a value proposition (e.g. (Adner 2006; Adner and Kapoor 2010; Ritala et al. 2013)). Ecosystems adds to the existing literature on innovation systems by emphasizing the complex network effects within the system and using an analogy from biology, we hope to better capture this complexity.

The broad-based interest, spanning multiple disciplines, have caused the knowledge base for ecosystems to become fragmented. The validity of ecosystems analogy to innovation systems has been critiqued where for example Oh et al. (2016) suggested that the concept of innovation ecosystem has begun to infiltrate spaces more traditionally described by such concepts as innovation system, triple-helix, or cluster which then this has led to ambiguous usage and application of the concept. A debate is undergoing on the validity and the value added of the ecosystem approach to innovation (Oh et al. 2016; Ritala and Almpanopoulou 2017). There appears to be no consensus about the progress made in the academic literature on the concept of ecosystems. The take-away from the debate is that too many unconnected concepts has made understanding of the concept difficult if not impossible.

In order to reduce the conceptual ambiguity of the concept of innovation ecosystem, this study uses GST as a lens to evaluate the discussion in the literature. Systems thinking is the cognitive process of studying and understanding systems of every kind. This framework is increasingly being used to tackle problems in a wide variety of disciplines, such as business, management and innovation studies (Jenkins and Youle 1968; Johannessen 2013; Mele et al. 2010). The systems approach in general assists in studying the functions of complex organizations and has been utilized as the base for the new kinds of approaches and therefore its terminology such as "ecosystem", has emerged to describe these organizations (Anderson 1999; Mele et al. 2010; Peltoniemi and Vuori 2004).

In this study, we analyze the ecosystem approach in the context of innovation studies in order to examine major system attributes discussed within the literature. The investigation is operationalized through GST and its five major attributes: components, interaction, dynamism, environment and goal. The quantitative analysis is complemented with an in-depth qualitative assessment and review of the literature. To guide the study, the following research question has been set:

- How has the systems approach and its attributes been adopted into innovation ecosystem literature?

The bibliometric analysis reveals the core body of knowledge in innovation ecosystem literature. The study identifies the major system attributes and further diagnoses the system attributes within the emerging literature of "innovation ecosystem". The main contributions of this paper are 1) A mixed-method approach applying GST to full-text science documents and 2) detecting the system attributes that have been discussed more than others have. These contributions are followed by discussions to identify fertile areas for future research.

The next section, reviews the traditional discussions on the systems and consequently ecosystem approach in business and management studies. The GST attributes are extracted from this literature review.

2. What is innovation ecosystem?

One of the trending prefixes to the ecosystem term within the context of Business & Management (BM) literature is "innovation ecosystems". Looking at the possible origins of the use of the ecosystem term, we can identify several factors that could explain its emergence and adoption. Innovation is an increasingly distributed and collective process involving a variety of components and interaction among them (Freeman and Luc 1997). Scholars have proposed the construct of innovation ecosystem to capture the multiplicity and complexity of the innovation process (see, e.g., Adner, 2006; Iansiti and Levien, 2004; Moore, 1993). *Ecosystem* in innovation studies context has been defined as the alignment structure of a multilateral set of partners that need to interact for a focal value proposition to materialize (Adner 2017). The term "ecosystem" was first utilized in the fields of business and economics by Rothschild (1990). In Rothschild's book "Bionomics", he is promoting understanding economics through how biological systems operate. Later on, Moore in (1993) took the ground and introduced the term "Business ecosystem" by which he emphasized the essentiality of competition among ecosystem components. Moore further stressed the dynamics that regenerate the interactions between organisms and the environment. Based on Suominen (2016), innovation ecosystem is a distinguished cluster, thriving by itself as a standalone domain among clusters such as *business ecosystem*, *knowledge ecosystem*, *digital ecosystem* and *platform*, but scholarly discourse is not set on this or any other definition of an innovation ecosystem or platform.

Terms platform and ecosystem are often seen to be connected even to the point that the terms are used interchangeably; however, there is an important distinction between the terms. Adner (2014) sees platforms as elements of a broader ecosystem, in which ecosystems are "communities of associated actors defined by their networks and platform affiliations". While ecosystems focus on structure and interdependence, platforms are concerned with governance. Platforms, he explains, "hold a hub position in a network of interactions" and "exercise power through centrality". Platforms also have an interest in the governance of interfaces, as well as their access, incentives, and control. Platforms can play the role of what Adner refers to as a "focal actor" in an ecosystem. Similarly, it has been argued that platforms often have a leader or a "keystone firm" which plays a central, orchestrating role within the platform's network or ecosystem (Gawer and Cusumano 2014; Iansiti and Levien 2004). Therefore, platform is conceptualized as a federation and coordination caused by a leadership to the constitutive agents who can innovate and compete; an ecosystem, on the other hand, suggests a more equitable relationship, specifically different pieces that work together to mutual benefits.

This paper is taking one step further into a detailed diagnosis of the contextual structure of extant literature to see how the ecosystem framework has been appropriated. To date, there are approximately 900 papers adopting the term ecosystem within B&M and organizational studies. An analysis of the highly referenced work within the community reveals a distinguished cluster of "innovation ecosystem". Yet, by reading the paper abstracts and the reference list, it is apparent that majority of the papers use the term ecosystem but are not referring to the scholarly

community as the reference point. The core of innovation ecosystem literature draws from the work of Moore (1993), Dhanaraj (2006) and Chesbrough (2003). The work highlight issues such as triple-helix and university-industry interaction and regional and sector-based innovation systems. Literature suggests that innovation ecosystems are distinct from innovation systems because they are not confined to a particular nation, geographic region, or industry. Nevertheless, the theoretical foundation, point of departure from existing literature, and even the definition of innovation ecosystems remains little understood as pointed out by Suominen (2016) and Hajikhani (2017), and subsequently form the focus of this study.

This paper focuses on novel approaches to enrich bibliometric data by the full text processing of the papers. Full text offers a new field of investigation, where the major problems arise around the concepts definitions and elaborations, which needs to be utilized in parallel with bibliometric methods to provide an accurate insight of the concept development. In the next section, the attributes of the holistic view are investigated by referring to traditional and classic discussions of systems in business and management studies. The general system theory attributes are extracted from these discussions in order to diagnose the growing literature on ecosystemic approach on a subdomain of innovation context.

3. System approach

A systems approach is commonly adopted in various disciplines because it offers a way to approach complex and persistent problems more effectively (Goodman 2015). Systems approach offers an interdisciplinary manner to explore systems in nature, in society and in many scientific domains as well as a framework with which we can investigate phenomena from a holistic approach (Capra 1997). Systems approach encompasses a wide field of research with different conceptualizations and areas of focus (e.g. Mele (2010) offers a brief review on system approach to B&M discipline).

Foundational to systems thinking was Von Bertalanffy's work on introduction of general systems theory (GST) in the late 1920s (Mele et al. 2010). Von Bertalanffy defines GST as "a logical-mathematical discipline, in itself purely formal but applicable to the various empirical sciences". For sciences concerned with 'organizational wholes,' GST would be of similar significance to probability theory for sciences concerned with 'chance events'; the latter, too, is a formal mathematical discipline which can be applied to diverse fields such as thermodynamics, biological and medical experimentation, genetics, life insurance statistics, etc. (Wieland et al. 2012). GST, like complexity theory, emerged from the realization that a reductionist view of the world, where objects and events could be understood in terms of their constituent parts and where these parts fit together like cogs in a machine, could not adequately capture the complexity of adaptive systems. (Bertalanffy 1969). Because GST utilizes a logical and formal approach to analyze and describe the phenomena under investigation, it can be applied to numerous scientific disciplines with distinct and disparate conceptualizations and areas of focus (Wieland et al. 2012).

A systems definition is very context-driven. Various organizations can define system architecture in different ways. Von Bertalanffy wanted to further identify the connection between business organizations and their environment, and to do so, he explored the relationships between employees, customers, and company output. His theories can best be illustrated by this quote from his compilation of articles titled General System Theory: Foundations, Development, and Applications:

"We may state as characteristic of modern science that this scheme of isolable units acting in one-way causality has proved to be insufficient. Hence the appearance, in all fields of science, of notions like wholeness, holistic, organismic, gestalt, etc., which all signify that, in the last resort, we must think in terms of systems of elements in mutual interaction." (Bertalanffy 1969)

What he means is that each element of a system must be looked at individually. Additionally, the result of the interactions must be explored, as well as the whole that is created as a result of those interactions (Saylor 2005).

A system can be defined as an entity that forms a coherent whole, with a boundary around it which functions both to distinguish internal and external elements and to identify input and output relating to and emerging from the entity (Mele et al. 2010). A systems theory is a theoretical perspective that analyzes a phenomenon seen as a whole and not as simply the sum of elementary parts, focusing on the interactions and on the relationships between parts in order to understand an entity's organization, function, and outcomes. This perspective implies a dialogue between holism and reductionism (Mele et al. 2010). One its most important characteristics is that it is composed of hierarchy of sub-systems. For example, the world can be considered-to be a system in which various national economies are sub-systems. Figure 1 represents the conceptual sketch of a system.

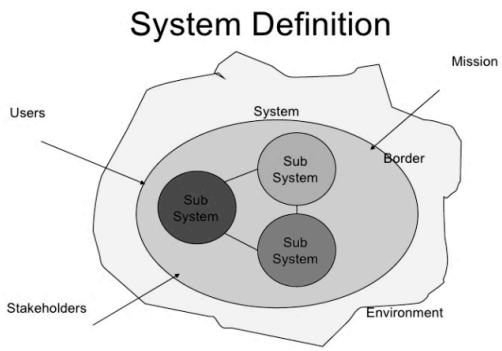


Figure 1. System Definition, adopted from Emmanuel Fuchs (2007)

Based on the reviewing the GST literature, we categorize the five major attributes of a system accordingly:

> Attribute #1 (Components): A system consists of interacting elements. It is set of inter-related and inter-dependent parts arranged in a manner that produces a unified whole. Other names to describe this attribute are: Actors, components, users, stakeholders (positions), and entities.
>
> Attribute #2 (Interaction): The various sub-systems should be studied in their inter-relationships rather than in isolation from each other. The emphasis is on relations, communication, and the process in which the system generates links, activities, and interactions to describe the interacting elements.
>
> Attribute #3 (Dynamism): A system does not exist in a vacuum. It receives information, material, and energy from other systems as inputs. These inputs undergo a transformation process within a system and leave the system as output to other systems. It has also been referred as the status of continuous change in the system.
>
> Attribute #4 (Environment): An organizational system has a boundary that determines which parts are internal and which are external. An organization is a dynamic system as it is responsive to its environment. It is vulnerable to change in its environment. This attribute considers the notion of the boundary/boarder of the system.
>
> Attribute #5 (Goal): A desired result or possible outcome that a system envisions to achieve is called the state of a goal in system literature. This attribute points out the purpose and mission of the system structure.

The extracted attributes can be seen in the seminal works on innovation ecosystems by Adner (2006 and 2010). Components has been explicitly referred as such (either complementors or intermediaries) among firms and end customers. Interaction is mentioned by the collaborative network and collaborative arrangements in the context of reducing the cost of a firm's coordination. Dynamism has been discussed as the evolutionary process within the ecosystem where a process takes place iteratively as assessment of the target for maximizing the value proposition of the ecosystem. Environment is accounted for by focused case studies. Finally, Goal is reflected in the context of the study of risk assessment and expectation alignment.

We suggests that system theory and its attributes can provide a powerful perspective and methodological lens for the analysis of the newly emerging ecosystem approach in innovation studies. Therefore, by introduction of system's main attributes, it is possible to examine and detect the attributes in the focus literature.

4. Methodology

The research is conducted through a mixed methods study that utilizes both qualitative and quantitative methods. A systematic mapping study was selected as the research methodology for this study. Bibliometric data analysis was utilized to collect the relevant literature. Automated full-text analytics, text segment identification, extraction and coding of articles combined with manual supervision to observe the discussion of system attributes.

4.1. Data retrieval and bibliometric analysis

Due to the vast volume of publications, methods for analyzing, organizing, and accessing information from large databases are in great need. Systematic reviews aim to address these problems by identifying, critically evaluating, and integrating the findings of all relevant, high-quality individual studies addressing one or more research questions (Baumeister and Leary. 1997; Cooper 2003). A modified version of the systematic mapping process described in Petersen (2008) was used for the study. We also use guidelines for a systematic literature review described by Kitchenham and Charters (2004) to search for relevant papers. Bibliometric data analysis was conducted to provide quantitative analysis of academic literature. Bibliometrics is known as statistical analysis of written publications and citation analysis (Nicolaisen 2010). Many research fields use bibliometric methods to explore the impact of their field, set of researchers, or a particular paper (Glänzel 2015). We use bibliometrics to explore the literature, which is relevant for further and qualitative content analysis. The bibliometric data collection and analysis was done with Network Analysis Interface for Literature Studies "NAILS"[1].

Web of Science (previously known as ISI Web of Knowledge) includes 90 million documents indexed and is considered to be one of the most important databases for scientific bibliometric data. The Web of Science (WoS) core collection was incorporated to enrich the coverage to all types of indexed documents.

The literature review reveals that the adoption of the ecosystem approach in innovation studies has been influenced by several complementary streams of research. In order to simplify the classification of the relevant literature, we used a combination of previously-identified supporting subject areas from WoS to consolidate and group the various publication outlets into five broad research streams (see Table 1).

Table 1. Functional discipline categories

[1] Network Analysis Interface for Literature Studies "NAILS" developed and published by 2015 Knutas et al. (Knutas et al. 2015) access from: http://nailsproject.net

Category	*Functional discipline (sub discipline)*
I	MANAGEMENT
II	BUSINESS
III	OPERATIONS RESEARCH MANAGEMENT SCIENCE
IV	SOCIAL SCIENCES INTERDISCIPLINARY
V	ECONOMICS

Bibliometric data retrieved from targeted database (Thomson Reuters Web of Science Core Collection), was utilized for building the search query with initial keywords. The query was built by using the keywords and boolean operators in order to be executed. Table 2 shows the searching query and the development date of the search.

Table 2. Searching query

Search words	*innovation AND ("ecosystem*" OR "eco-system*" OR "eco system")*
Development date	Jun 2017
Databases	Web of knowledge (core collection)

Due to the specificity of the search query, the results has been fitted in the particular domain of innovation studies. The search resulted in 136 records. For these results an initial refining (including and excluding of the records) was done. This was done by retrieving the full-text version of the articles. The articles were retrieved using the WoS linkages to publisher's databases. In cases where online access was not available, the paper was obtained through the scholar's online research profiles (e.g. Research Gate). The full text of each article was reviewed by the author to eliminate those articles that did not meet the selection criteria. The main exclusion criteria for articles was that the publication needs to refer to innovation ecosystem directly in the title. This draws from the assumption that by having innovation ecosystem in the title we can be certain that the articles selected sufficiently describes and captures the breadth of potential topics associated with the approach of ecosystem in innovation studies. The selection process led to of the final dataset having 109 records data was imported to NAILS for in-depth analysis.

A general analysis of the data shows the exponential rise in the count of publications in recent years, which reflects increased interest in the topic. Figure 2 shows both yearly volume and accumulative measures of number of publications.

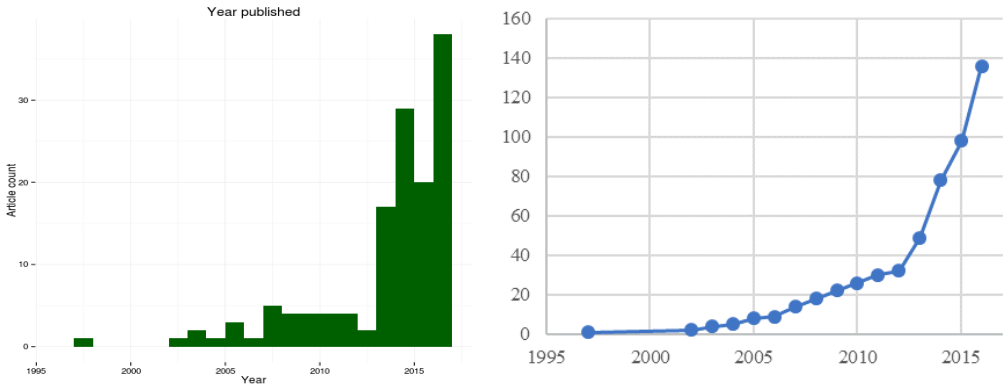


Figure 2. Publication volume

The NAILS analysis tool automatically calculates several metrics, such as PageRank (counts of the number and quality of links to a paper to determine a rough estimate of its importance) and In-Degree (the number of citations to a paper in a directed graph). A tailor-made report of these metrics was generated that provides an abstract/keyword analysis, productive authors/journals, and recommendations for relevant publications based on the citation data (Link to the NAILS report). In the Findings section below, analysis of the 109 records' bibliometric data is presented to document the core literature on the topic of "innovation ecosystem". A full content analysis was performed to the selected core literature for deeper review. A coding system was implemented to specify the use of system attributes defined in the previous section.

4.2. System attributes in literature

Text mining methods have benefited the traditional analysis of literature immensely (Basole et al. 2013; Bragge et al. 2007; Porter et al. 2002). Traditional literature analysis may be compromised by unintentional and intentional bias in the selection, interpretation, and organization of content, while the use of text mining strengthens the discovery and insights, especially from large data sources (Delen and Crossland 2007). Text analytics, also known as text mining, overcame these problems to a large extent by adding depth and intelligence to our ability to utilize a growing mass of unstructured text. However, the machine is incapable of understanding the context of a text; therefore a mixed method is required with human attention to enrich the analysis. Therefore, the methodological approach of this paper is to leverage text analytics in a mixed setup of quantitative and qualitative analysis of the literature.

Identification of system attributes in the selected literature was facilitated by automatic text segment extraction. Lexical queries were constructed based on each system attribute's definition as rules that specify what text segment should be identified and processed. Lexical queries organize the decision logic of identification and extraction of relevant text segments and combinations of text segments. Examples include identification of two or more terms in pre-

defined word proximity of each other. For example, if the interest is to understand the level of comprehension of a systems approach and accordingly the discussion on the matter of system attributes within the articles, this scenario would require the identification of factors within one paragraph proximity to the keywords related to system attribute #1 (actors and/or its equivalents) as well as to the keywords related to other four system attributes. Searching rules are operationalized using a string of boolean logic statements, shown in Table 3.

Table 3. System attributes searching queries

System attributes	*Query*	*Example*
Attribute #1	actor*, component*, user*, stakeholder*, entit*, position*, element*, part*	... is a challenging issue as participants leave and join the network…
Attribute #2	relation*, communication, link*, activit*, interact*, interrelated, inter-related, interdependent, inter-dependent, inter-relationship*, interrelationship*, interdependent, inter-dependent*, “dynamic system", dynamism, dynamic	... the graph corresponds to an interaction between the participants who are…
Attribute #3	input*, transformation, output*	... process, rather than an output, which always involves the participation…
Attribute #4	boundary, internal, external, responsive, environment, boarder*, sub-system*, structure	... boundaries, structure, and dynamics arising from relationships…
Attribute #5	purpose, mission, evolution*, co-evolution*	... calling for understanding of co-evolutionary processes across systemic levels in…

The analysis was performed using the MAXQDA software package (MAXQDA, 2017). Search queries were used to identify the text segments. MAXQDA analyzed the literature in response to the search query and rules, then generated text segments, identified co-occurrence of text segments, and provided insights on coded segments’ occurrence patterns. In total, 500 coded sections were randomly chosen for manual inspection to ensure accuracy. In order to estimate the overall accuracy of the system’s attributes text segments detection, we calculate the Recall and Precision as the commonly used evaluation measures. The mentioned measures challenge the accuracy of the selection of relevant text segments. The precision means how many detected text segments regarding the system attribute query are relevant and the recall looks for how many relevant text segments regarding the system attributes query has been selected. In our case, the system attributes searching query precision is 82% while its recall is 87%. There is no standard measures for this specific study context. In simple terms, the high precision ratio

means that the system attribute search query returned substantially more relevant results than irrelevant ones, while high recall means that the system attribute search query returned most of the relevant results.

5. Findings

The study of systems approach adoption to ecosystem literature was a primary topic of interest in this research. Due to the emerging nature of the ecosystem concept in innovation studies, the publication venues are disparate and it is difficult to identify a publication host for the concept development, discussion, and debates. Figure 3 shows the frequency of the most productive (quantity of publications) and popular (Sorted by higher number of received citations) publication venues and authors.

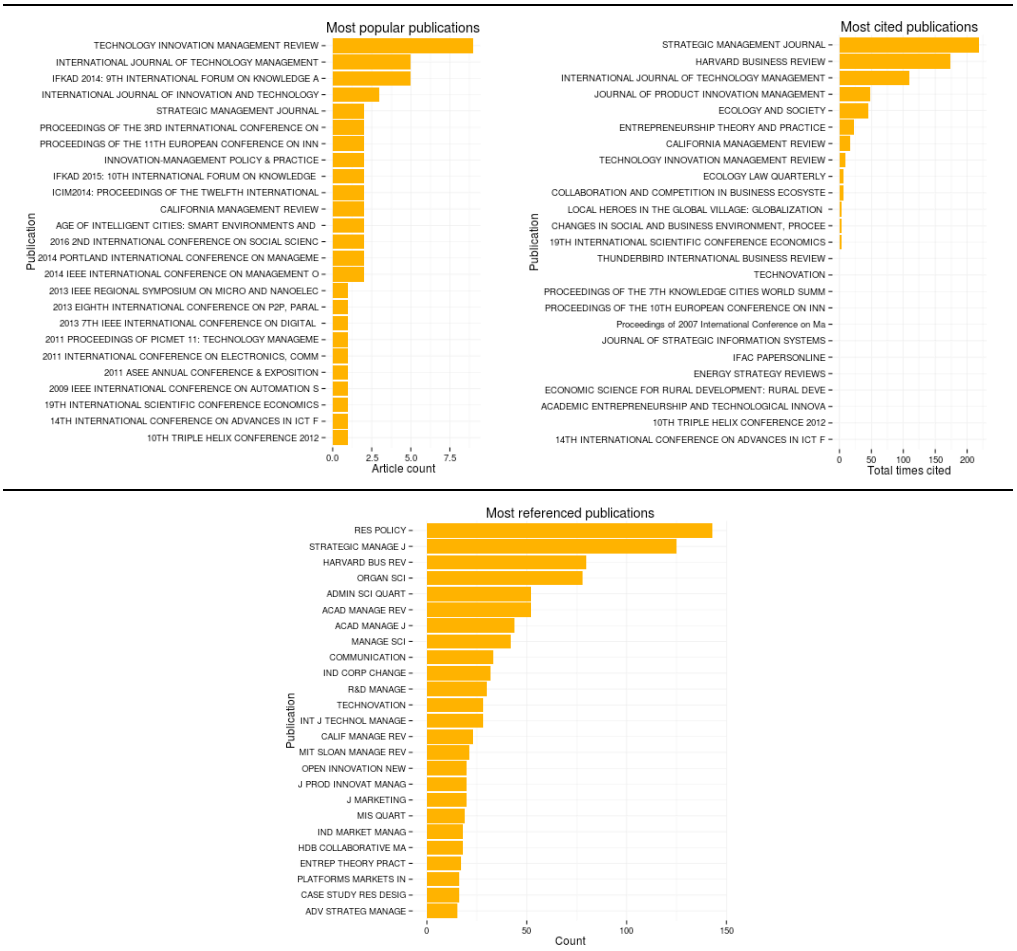


Figure 3. Frequency of most productive/popular publication venues and authors

The most popular publication venues have been “Technology and Management Review”, “International Journal of Technology Management” and “International Forum on Knowledge Management” while the most cited publications were “Strategic Management Journal”, “Harvard Business Review” and “International Journal of Technology Management”. Although journals are among the top publications, which focus on innovation ecosystem literature, still over two-thirds of all innovation ecosystem studies were published as part of a conference proceedings. This perhaps suggests that ecosystem concept has not yet been adopted as a core topic of interest in the innovation literature. Additionally, most referenced articles among the innovation ecosystem literature has been published in “Journal of Research Policy”, “Strategic Management Journal”, “Harvard Business Review” and “Organizational Science”.

Citation patterns are relevant to see which publications have been influential on the discussion and help to discover the patterns of knowledge diffusion. The network consists of nodes, which represent papers and edges which represent citations. As shown in Appendix 1, the size of the nodes indicates the in-degree value (the mean amount of connections per node on the graph) and an edge shows a citation between nodes.
Appendix 1 illustrates the evolutionary path of “innovation ecosystem” literature, taking the work of Adner as a point of reference. The network visualization illustrates the formation of the core literature of innovation ecosystem and the periphery. The first phase shows the works having innovation ecosystem as the terminology in their titles; the size of the nodes indicates the cumulative citation counts of the nodes. The work of Adner (2006) is visible as a dark node. The grey nodes are representing articles discussing innovation ecosystem related topics, which appeared before Adner’s (2006) work. The red nodes are the ones appearing after Adner’s (2006) work.
The evolutionary citation network reveals that among the innovation ecosystem related papers only 10% of publications are aligned with the core of innovation ecosystems literature. The other 90% approach innovation ecosystems from many different points of departure. Analysis of the textual content of the 90 % reveals interesting information. Big data and ICT emerge as central terms among 73% of the records. This relationship shows the usage of ecosystem to describe applications of data analytics in innovation studies. The term ecosystem has been considered in order to give introduction and unveiling studies, which incorporate big data analytics. Meanwhile over time the popularity of the terminology caused more framing and solid definition which a core and periphery is visible in the recent years.

This study focuses on the ones, which are connected and form a cluster base on citations. A seminal discussion by Adner on the topic of innovation ecosystem started in 2006 and after there has been a stream of literature on the topic. We focus on the studies, which are central to the discussions. The centrality to the discussion is based on citations and set by the in degree of publication. Core literature was also expected to have citations to the minimum two of the papers in the initial collection. This creates a core literature 13 publications. Table 4 illustrates the core set of publications on the discussion of innovation ecosystem.

Table 4. "Innovation ecosystem" core literature

Year Published	Authors Full Name	Document Title	Publication Name	Document Type	Times Cited	In-Degree
2010	ADNER, RON; KAPOOR, RAHUL	VALUE CREATION IN INNOVATION ECOSYSTEMS: HOW THE STRUCTURE OF TECHNOLOGICAL INTERDEPENDENCE AFFECTS FIRM PERFORMANCE IN NEW TECHNOLOGY GENERATIONS	STRATEGIC MANAGEMENT JOURNAL	Article	222	23
2006	ADNER, R	MATCH YOUR INNOVATION STRATEGY TO YOUR INNOVATION ECOSYSTEM	HARVARD BUSINESS REVIEW	Article	179	19
2014	GAWER, ANNABELLE; CUSUMANO, MICHAEL A.	INDUSTRY PLATFORMS AND ECOSYSTEM INNOVATION	JOURNAL OF PRODUCT INNOVATION MANAGEMENT	Article	53	3
2013	NAMBISAN, SATISH; BARON, ROBERT A.	ENTREPRENEURSHIP IN INNOVATION ECOSYSTEMS: ENTREPRENEURS SELF-REGULATORY PROCESSES AND THEIR IMPLICATIONS FOR NEW VENTURE SUCCESS	ENTREPRENEURSHIP THEORY AND PRACTICE	Article	22	5
2013	RITALA, PAAVO; AGOURIDAS, VASSILIS; ASSIMAKOPOULOS, DIMITRIS; GIES, OTTO	VALUE CREATION AND CAPTURE MECHANISMS IN INNOVATION ECOSYSTEMS: A COMPARATIVE CASE STUDY	INTERNATIONAL JOURNAL OF TECHNOLOGY MANAGEMENT	Article	11	2
2013	LETEN, BART; VANHAVERBEKE, WIM; ROIJAKKERS, NADINE; CLERIX, ANDRE; VAN HELLEPUTTE, JOHAN	IP MODELS TO ORCHESTRATE INNOVATION ECOSYSTEMS: IMEC, A PUBLIC RESEARCH INSTITUTE IN NANO-ELECTRONICS	CALIFORNIA MANAGEMENT REVIEW	Article	9	1
2014	FRENKEL, A; MAITAL, S	MAPPING NATIONAL INNOVATION ECOSYSTEMS: FOUNDATIONS FOR POLICY CONSENSUS	MAPPING NATIONAL INNOVATION ECOSYSTEMS: FOUNDATIONS FOR POLICY CONSENSUS	Book	7	2
2014	STILL, KAISA; HUHTAMAKI, JUKKA; RUSSELL, MARTHA G.; RUBENS, NEIL	INSIGHTS FOR ORCHESTRATING INNOVATION ECOSYSTEMS: THE CASE OF EIT ICT LABS AND DATA-DRIVEN NETWORK VISUALISATIONS	INTERNATIONAL JOURNAL OF TECHNOLOGY MANAGEMENT	Article	6	1
2013	BRUSONI, STEFANO; PRENCIPE, ANDREA	THE ORGANIZATION OF INNOVATION IN ECOSYSTEMS: PROBLEM FRAMING, PROBLEM SOLVING, AND PATTERNS OF COUPLING	COLLABORATION AND COMPETITION IN BUSINESS ECOSYSTEMS	Article; Book Chapter	6	1
2015	GASTALDI, LUCA; APPIO, FRANCESCO PAOLO; MARTINI, ANTONELLA; CORSO, MARIANO	ACADEMICS AS ORCHESTRATORS OF CONTINUOUS INNOVATION ECOSYSTEMS: TOWARDS A FOURTH GENERATION OF CI INITIATIVES	INTERNATIONAL JOURNAL OF TECHNOLOGY MANAGEMENT	Article	3	1

2014	JUCEVICIUS, GIEDRIUS; GRUMADAITE, KRISTINA	SMART DEVELOPMENT OF INNOVATION ECOSYSTEM	19TH INTERNATIONAL SCIENTIFIC CONFERENCE ECONOMICS AND MANAGEMENT 2014 (ICEM-2014)	Proceedings Paper	3	1
2016	PELLIKKA, JARKKO; ALI-VEHMAS, TIMO	MANAGING INNOVATION ECOSYSTEMS TO CREATE AND CAPTURE VALUE IN ICT INDUSTRIES	TECHNOLOGY INNOVATION MANAGEMENT REVIEW	Article	2	1
2016	VIITANEN, JUKKA	PROFILING REGIONAL INNOVATION ECOSYSTEMS AS FUNCTIONAL COLLABORATIVE SYSTEMS: THE CASE OF CAMBRIDGE	TECHNOLOGY INNOVATION MANAGEMENT REVIEW	Article	2	1

The attributes were categorized using the system definition in Section 3 and the operationalization of the analysis was explained in Section 4.2. The automated process of detecting the textual parts discussion any of the system attributes was facilitated with vocabulary affiliated with each distinguished system attribute. The process was followed by a manual review of the detected textual segments and confirming the assignments of the system attributes. Figure 5 is a cross tabulation which illustrates the classification of the system's attributes discussed within the papers.

	Attribute #1 (Components	**Attribute #2 (Interaction)**	**Attribute #3 (Dynamism)**	**Attribute #4 (Environment**	**Attribute #5 (Goal)**
(Gawer and Cusumano, 2014)	118	36	0	57	17
(Pellikka and Ali-Vehmas, 2016)	32	33	1	22	3
(Viitanen, 2016)	420	157	14	182	68
(Leten et al., 2013)	110	9	3	19	6
(Adner, 2006)	59	6	2	15	0
(Adner and Kapoor, 2010)	143	69	4	68	12
(Jucevičius and Grumadaitė,	21	23	8	11	5
(Gastaldi et al., 2013)	33	54	9	24	9
(Brusoni and Prencipe, 2013)	72	72	1	116	15
(Nambisan and Baron, 2013)	56	57	1	41	1
(Ritala et al., 2013)	133	64	0	45	18
(Still et al., 2014)	71	111	27	30	11

Figure 5 Distribution of system attributes in the core literature. (Calculation of the symbol size refers to each publication)

The co-occurrence of codes for system attributes indicates the inter-relatedness and inter-dependence of system attributes discussion among literatures. The following diagram (Figure 6) is a code co-occurrence model, which displays the system attributes text segments overlapping in the studied literature. The system attributes are shown as labels with their respective names and a line to represent the co-occurrence of the codes in a document.

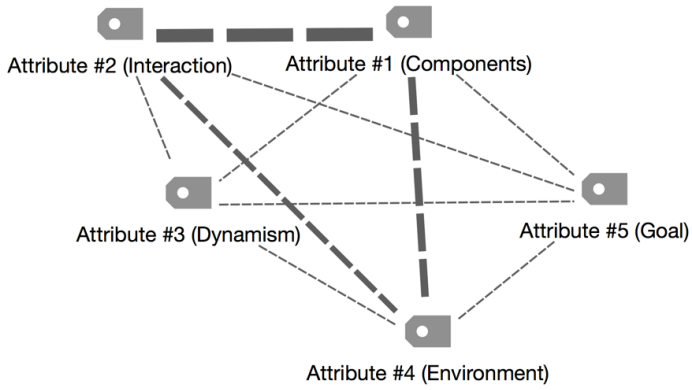


Figure 6. System attributes text segments co-occurrence model.

The co-occurrence of system attribute #1, #2 and #4 indicates that the discussions of *components, environment* and *interaction* are happening interchangeably within a closer proximity. This implies that the notion of introducing components of an ecosystem and discussing the interaction among them has been fairly established and the description of environment attribute co-exists strongly with the other two attributes. Conversely, the discussion of attributes associated to dynamism and goal of the system has not occurred with the same strength in the literature.

6. Discussion

The concept of ecosystem in business and strategy has been extensively discussed within the recent years. This discussion centered on the notion that the biological analogy of a system with different components being linked together through nutrient cycles and energy flows. Although analogies play a crucial role in human comprehension of complex systems, we need to understand the analogy made to understand its usefulness and applicability. This study has made an effort to highlight how the ecosystem term is used in current literature. Particularly this identifies what part of the biological analogy is useful for the business and strategy literature.

The study created, through an automatic text segment extraction analysis, five system attributes, which all were visible in various portions among the studied literature. In making these attributes explicit, the study adds to the current understanding by highlighting the type of nutrients and energy flows we look towards while using the analogy. The study provided a comprehensive classification and analysis business and strategy related ecosystem literature. In total, 109 articles identified from all peer reviewed publication venues over the past two decades from WoS core collection. The bibliometric data of the publications was utilized to extract 13 publications as the core set of publications, which are the foundational on the topic of innovation ecosystem.

The results highlight why we have adopted the term ecosystem. The core literature uses the term to focus in particular to components, their interaction in a given environment. This seems to be the motivation of talking about innovation ecosystem – using the terminological shift to argue that we would need to look more thoroughly on the complexities of actors and actor roles within a value production system. In Adner's (2006) work, the core thesis is that within an ecosystem components (firms) through interaction are able to create value that no single actor in the system is capable to create. The additionality of the ecosystem is totally embedded on the components and their interactions, in particular thought he complementing nature of the components (James F Moore 1996). Similarly, the environment is front and center to the debate. Moore (1996) writes about the co-evolution of the environment and that their needs to be a shared understanding on the environment (Iansiti & Levien 2004) the components are working in.

The results encourage more attention to be put on the *Dynamism* and *Goal* system attributes. Dynamism as a system attribute was intended to be a holistic glance on solutions which results in collaborative innovation that recognizes the role of inputs. The outputs of a component are considered as an initiative input for another component. The importance of collective input and assessing expectations in order to evaluate the ecosystem risk was taken by appreciating a group risk-assessment process that will deliver better expectations and more relevant strategy (Adner 2006). Adner (2010) perceived the notion of dynamism as the process that the outputs of upstream suppliers serve as inputs to the focal actor. Gawer (2014) conceptualized dynamism as transformation and evolution to a certain stage caused by forces such as network effects and competition. The notion of goal as a system attribute considers the whole system structure as a set to produce and justify a certain goal. This is often called the purpose, outcome, or goal of a system. The presentation of a system goal occurs variably in each individual paper. Most prominent the goal attribute is in case studies, such as Wang et al. (2014) for smart cities or Rong et al. (2013) on optimizing a decision-making process within the public sector. As dynamism and goal are seen in the current literature, they imply towards case studies with particular dynamic behavior and share goal. However, at a theoretical level a shared goal and somewhat implicit dynamic behavior among components of the system should be expected. Within the business and strategy domain, we can expect that actors make strategic decision to act in an ecosystem. Actors' capability to manage their position and that of other in the ecosystem governs their ability to create and capture value from participation (Hannah 2013). Current literature, however, suggests that actors operate towards their own goals and that there is now shared understanding of dynamism or goals with an ecosystem (Jacobides et al. 2014). This assumption requires more attention, as being able to change and share a goal seem central to ecosystem health - a key indicator of "well-being, longevity and performance" of the ecosystem (Hyrynsalmi et al. 2012).

GST has been advanced from conceptual ground to a more solid toolkit set. System dynamics has adopted an engineering approach and introduced a set of toolkits to capture and materialize the system approach. System dynamics offers an approach to understanding

the nonlinear behavior of complex systems over time using a set of constructs such as stocks, flows, internal feedback loops, table functions, and time delays (Sterman 2000). Also, an attempt has been made to bridge the complexity of a system by network visualization (Russell et al. 2011). Network visualization and social network metrics adoption have benefited the community to document the ecosystem evolution. Network visualization can provide evidence about ecosystem transformation and opportunities for orchestrating this transformation (Still et al. 2014). Social Network Analysis has been utilized for better understanding of ecosystem transformation. Visual representation of social networks is important to understanding the network data and conveying the result of the analysis (Huhtamäki and Rubens 2016). However, the limitation with capability of handling the data and various properties of nodes make the method incapable of showing the full transformation of outcomes from a certain stage of a system to another stage.

To conclude, the results provide interesting implications to further emphasize system component usage in innovation ecosystem literature. The work invites other scholars to contribute to the evolution and development of the systemic nature of the ecosystem concept within innovation studies. We need to employ models and tools that can simplify the complexity of social and economic exchange in a meaningful way without eliminating the richness that a solid traditional system approach provides. The resources for these models and tools can be found through a collaborative effort from diverse academic disciplines to provide the cross-fertilization that is needed for the next step of analysis.

From a theoretical perspective, this paper contributes to the overall understanding of the ecosystem approach adoption in innovation literature and the areas where the approach should be improved or further studied. From a methodological perspective, this study shows the design, applicability, and value of text analytics and citation analysis for concept evolution, knowledge discovery, and literature reviews, and is one of few attempts to do so in the broad set of literature in the field. Studies such as this provide a quantitative analysis of the state of the art as a complement to traditional qualitative methods of reviewing the literature. They can be used as a tool to identify the authors, documents, and journals most widely read among the researchers in a given discipline and also to detect relational links between them as a basis for more qualitative reviews.

References

Adner, R. (2006). Match Your Innovation Strategy to Your Innovation Ecosystem Match Your Innovation Strategy to Your Innovation Ecosystem. *Harvard business review*.

Adner, R. (2017). Ecosystem as Structure: An Actionable Construct for Strategy. *Journal of Management*, *43*(1), 39–58. doi:10.1177/0149206316678451

Adner, R., & Euchner, J. (2014). Innovation Ecosystems: An Interview with Ron Adner. *Research Technology Management*, *57*(6), 10–14. doi:10.5437/08956308x5706003

Adner, R., & Kapoor, R. (2010). Value creation in innovation ecosystems: how the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal*, *31*(3), 306–333. doi:10.1002/smj.821

Anderson, P. (1999). Complexity theory and organizational science. *Organizational science*, *10*(3), 216–232.

Basole, R. C., Seuss, C. D., & Rouse, W. B. (2013). IT innovation adoption by enterprises: Knowledge discovery through text analytics. *Decision Support Systems*, *54*(2), 1044–1054. doi:10.1016/j.dss.2012.10.029

Baumeister, R. F., & Leary., M. R. (1997). Writing narrative literature reviews. *Review of General Psychology*, *24*(4), 230–235. doi:1089-2680/97153.00

Bertalanffy, K. L. von. (1969). General System Theory, 296.

Bornmann, L., & Marx, W. (2014). The wisdom of citing scientists. *Journal of the Association for Information Science and Technology*, *65*(6), 1288–1292. doi:10.1002/asi.23100

Bragge, J., Relander, S., Sunikka, A., & Mannonen, P. (2007). Enriching literature reviews with computer-assisted research mining. Case: Profiling group support systems research. *Proceedings of the Annual Hawaii International Conference on System Sciences*. doi:10.1109/HICSS.2007.209

Braun, T., Bujdosó, E., & Schubert, A. (1987). *Literature of analytical chemistry: a scientometric evaluation*. *CRC Press*.

Capra, F. (1997). The web of life. *New York: Doubleday-Anchor Book*.

Cooper, H. (2003). Psychological Bulletin: Editorial. *Psychological Bulletin*, *129*(1), 3–9. doi:10.1037/0033-2909.129.1.3

Delen, D., & Crossland, M. D. (2007). Seeding the survey and analysis of research literature with text mining. *Expert Systems with Applications*, *34*(3), 1707. http://www.sciencedirect.com/science/article/pii/S0957417407000486

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of medical Internet research*, *13*(4). doi:10.2196/jmir.2041

Freeman, C., & Luc, S. (1997). *The economics of industrial innovation*. Psychology Press.

Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, *43*(7), 1239–1249. doi:10.1016/j.respol.2014.03.006

Gawer, A., & Cusumano, M. A. (2014). Industry platforms and ecosystem innovation. *Journal of Product Innovation Management*, *31*(3), 417–433. doi:10.1111/jpim.12105

Glänzel, W. (2015). Bibliometrics-aided retrieval: where information retrieval meets scientometrics. *Scientometrics*, *102*(3), 2215–2222. doi:10.1007/s11192-014-1480-7

Goodman, M. (2015). Importance of Systems Thinking Today. *appliedsystemsthinking*. http://www.appliedsystemsthinking.com/importance.html. Accessed 20 February 2017

Gorraiz, J., Gumpenberger, C., & Schlögl, C. (2014). Usage versus citation behaviours in four subject areas. *Scientometrics*, *101*(2), 1077–1095. doi:10.1007/s11192-014-1271-1

Hajikhani, A. (2017). Emergence and dissemination of ecosystem concept in innovation studies : A systematic literature review study. In *Hawaii International Conference on System Sciences (HICSS) 2017* (pp. 1–12). http://hdl.handle.net/10125/41796

Hannah, D. P. (2013). Puzzles or Pieces: Competition in Nascent System Industries. *Academy of Management Proceedings*, *2013*(1), 14988–14988. doi:10.5465/AMBPP.2013.14988abstract

Huhtamäki, J., & Rubens, N. (2016). Exploring Innovation Ecosystems as Networks: Four European Cases. *Proceedings of the Hawaii International Conference on System Sciences HICSS-49: January 5-8, 2016, Grand Hyatt, Kauai*, 10. doi:10.1109/HICSS.2016.560

Hyrynsalmi, S., Mäkilä, T., Järvi, A., Suominen, A., Seppänen, M., & Knuutila, T. (2012). App store, marketplace, play! an analysis of multi-homing in mobile software ecosystems. *CEUR Workshop Proceedings*, *879*, 59–72. doi:10.1007/s00199-006-0114-6

Iansiti, M., & Levien, R. (2004). Keystones and dominators: framing operating and technology strategy in a business ecosystem. *Harvard Business School, Working Paper*, 3–61.

Jacobides, M., Veloso, F., & Wolter, C. (2014). *Ripples through the value chain and positional bottlenecks: Innovation and profit evolution in a competitive setting*. London.

Jenkins, G. M., & Youle, P. V. (1968). A Systems Approach to Management. *OR*, *19*, 5. doi:10.2307/3007468

Johannessen, J. a. (2013). Innovation: a systemic perspective - developing a systemic innovation theory. *Kybernetes*, *42*(8), 1195–1217. doi:10.1108/k-04-2013-0069

Jonkers, K., Derrick, G. E., Lopez-Illescas, C., & Van den Besselaar, P. (2014). Measuring the scientific impact of e-research infrastructures: a citation based approach? *Scientometrics*, *101*(2), 1179–1194. doi:10.1007/s11192-014-1411-7

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, *33*(TR/SE-0401), 28. doi:10.1.1.122.3308

Knutas, A., Hajikhani, A., Salminen, J., Ikonen, J., & Porras, J. (2015). Cloud-based Bibliometric Analysis Service for Systematic Mapping Studies. In *Proceedings of the 16th International Conference on Computer Systems and Technologies* (Vol. 1008, pp. 184–191). doi:10.1145/2812428.2812442

Mele, C., Pels, J., & Polese, F. (2010). A Brief Review of Systems Theories and Their Managerial Applications. *Service Science*, *2*(1–2), 126–135. doi:10.1287/serv.2.1_2.126

Moore, J. F. (1993). Predators and prey: a new ecology of competition. *Harvard Business Review*, *71*(3), 75–86. doi:Article

Moore, J. F. (1996). The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems. *Leadership*, 297. doi:10.1017/CBO9781107415324.004

Nicolaisen, J. (2010). Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics. *Journal of the American Society for Information Science and Technology*, *61*(1), 205–207.

Oh, D.-S., Phillips, F., Park, S., & Lee, E. (2016). Innovation ecosystems: A critical examination. *Technovation*. doi:10.1016/j.technovation.2016.02.004

Peltoniemi, M., & Vuori, E. (2004). Business ecosystem as the new approach to complex adaptive business environments. *Proceedings of eBusiness Research Forum*, 267–281.

Petersen, K., Feldt, R., Mujtaba, S., & Mattson, M. (2008). Systematic Mapping Studies in Software Engineering. *12th International Conference on Evaluation and Assessment in Software Engineering (EASE 2008)*, 71–80.

Porter, A. L., Kongthon, A., & Lu, J. C. (2002). Research profiling: Improving the literature review. *Scientometrics*, *53*(3), 351–370. doi:10.1023/A:1014873029258

Ritala, P., Agouridas, V., Assimakopoulos, D., & Gies, O. (2013). Value creation and capture mechanisms in innovation ecosystems: a comparative case study. *International Journal of Technology Management*, *63*(xxxx), 244–267. doi:10.1504/IJTM.2013.056900

Ritala, P., & Almpanopoulou, A. (2017). In defense of "eco" in innovation ecosystem☆. *Technovation*, (January), 4–7. doi:10.1016/j.technovation.2017.01.004

Rong, K., Shi, Y., & Yu, J. (2013). Nurturing business ecosystems to deal with industry uncertainties. *Industrial Management & Data Systems*, *113*(3), 385–402. doi:10.1108/02635571311312677

Rothschild, M. (1990). *Bionomics: Economy As Ecosystem*. http://www.amazon.com/Bionomics-Economy-Ecosystem-Michael-Rothschild/dp/0805019790

Russell, M. G., Still, K., Huhtamäki, J., Yu, C., & Rubens, N. (2011). Transforming Innovation Ecosystems through Shared Vision and Network Orchestration. *Triple Helix IX International Conference: Silicon Valley: Global Model or Unique Anomaly?*, 1–21. doi:10.1017/CBO9781107415324.004

Saylor. (2005). Historical and Contemporary Theories of Management, 1–7.

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. *Management*. doi:10.1057/palgrave.jors.2601336

Still, K., Huhtamäki, J., Russell, M. G., & Rubens, N. (2014). Insights for orchestrating innovation ecosystems: the case of EIT ICT Labs and data-driven network visualisations. *International Journal of Technology Management*, *66*(2/3), 243. doi:10.1504/IJTM.2014.064606

Suominen, A., & Seppänen, M. (2016). Innovation Systems and Ecosystems: a Review and Synthesis. In *ISPIM 2016*.

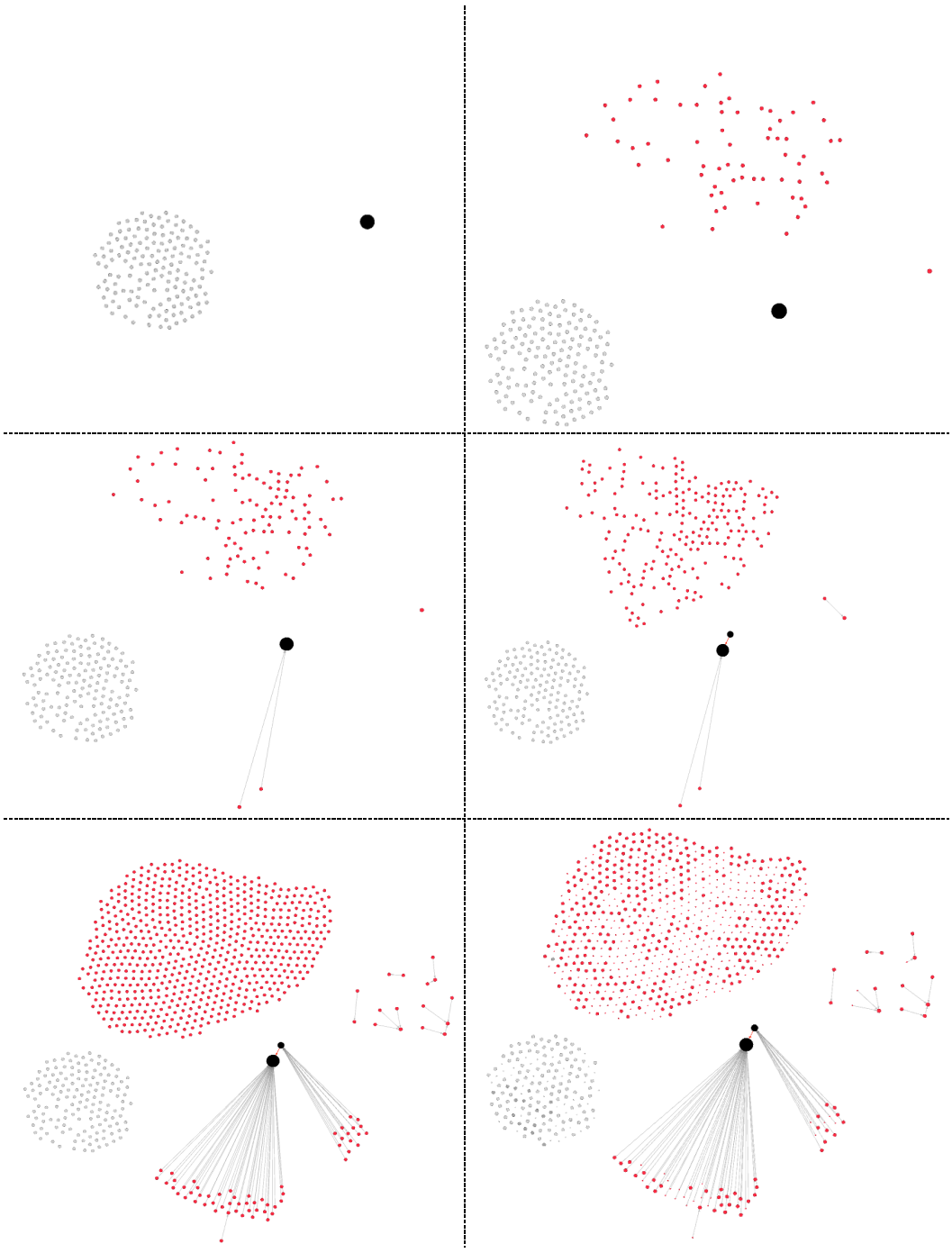
Vinkler, P. (2010). The Evaluation of Research by Scientometric Indicators. *The Evaluation of Research by Scientometric Indicators*, 1–313. doi:10.1533/9781780630250

Waltman, L., Van Eck, N. J., & Wouters, P. (2013). Counting publications and citations: Is more always better? *Journal of Informetrics*, *7*(3), 635–641. doi:10.1016/j.joi.2013.04.001

Wang, K., Chen, J., & Zheng, Z. (2014). Insigma's Technological Innovation Ecosystem for Implementing the Strategy of Green Smart City. *PICMET 2014 - Portland International Center for Management of Engineering and Technology, Proceedings: Infrastructure and Service Integration*, 892–899. http://www.scopus.com/inward/record.url?eid=2-s2.0-84910140089&partnerID=tZOtx3y1

Wieland, H., Polese, F., Vargo, S. L., & Lusch, R. F. (2012). Toward a Service (Eco)Systems Perspective on Value Creation. *International Journal of Service Science, Management, Engineering, and Technology*, *3*(3), 12–25. doi:10.4018/jssmet.2012070102

Appendix 1



Publication III

Hajikhani, A., Porras, J., Melkas, H.

Brand Analysis in Social Network Services: Results from Content Analysis in Twitter Regarding the US Smartphone Market

Reprinted with permission
International Journal of Innovation and Technology Management
Vol 14, Issue 02. 2017
doi: 10.1142/S0219877017400089

Brand Analysis in Social Network Services: Results from Content Analysis in Twitter Regarding the US Smartphone Market, Hajikhani A., Porras J., Melkas H., International Journal of Innovation and Technology Management, Volume 13, Issue 6, 2016, 32] [DOI: 10.1142/S0219877017400089]

Brand Analysis in Social Network Services: Results from Content Analysis in Twitter Regarding the US Smartphone Market

Arash Hajikhani*, Jari Porras, Helinä Melkas

School of Business and Management
Lappeenranta University of Technology
Skinnarilankatu 34, FI-53851, Lappeenranta, Finland†
arash.hajikhani@lut.fi‡
jari.porras@lut.fi
helina.melkas@lut.fi

Received 17 April 2015
Revised 31 October 2015
Accepted 5 November 2015
Published

Marketing is the front end strategy of firms to communicate the value of their product or services to customers; therefore, innovations in marketing have tremendous value in comparison to the whole innovation strategy of firms. The emergence of social network services (SNSs) as a dominant communication platform among firms and users provides an opportunity to evaluate the innovativeness of a firm's marketing strategy. With an analysis of twitter data, the study indicates how users react to content from different profile types. This result could inspire firms and the social media strategists of companies to diversify their content over multiple user profiles.

Keywords: Social network services; twitter; sentiment analysis; crowd intelligence evaluation; brand presence.

1. Introduction

If the first Internet revolution was its wide adoption as a personal and business platform, then the second Internet revolution has undoubtedly been the recent explosion in technologies affiliated with Social Networking Services (SNSs). SNSs can be described as online tools to share content, opinions, perspectives and insights. Users can create content, or merely observe and disseminate material [Holtzblatt (2011)].

Records show a remarked yearly increase in new SNS users, which will only accelerate because of the growing popularity of social media among all demographics. Social network services such as Facebook and Twitter are quite dominant today and has users worldwide. These technologies have changed how people lead their social lives, fundamentally and

* Corresponding author

altered marketing and communications. Eventually, users' trust is built and forming in SNSs, and proof of that is the level of trust among people and content in social media regarding their health and wellness issues [Elkin (2008)]. Moreover, social network services has shown an enormous impact on how private-sector businesses operate. The essence of the SNSs is the relationships and connections they enable between people and organizations [Nair (2011)]

The necessity of moving toward social media is obvious; companies are entering these areas because they are immensely popular and thus an economical and highly efficient way to reach large audiences [Kaplan and Haenlein (2010)]. Earlier studies also prove that many major businesses have already moved to this area of communication [Stelzner (2013)]. Eventually, with the emergence of SNSs another front line is created for challenging companies to be innovative in channels which they maintain in social media.

The wisdom of the crowd, documented on the web, is apparently playing a key role in our major decisions in almost any context [Nair (2011)]. Therefore, investigating a practical way of how to explore and mine valuable information from user-generated data will be important. Particularly User-Generated Content (UGC) in social media could be capable of capturing some features such as the public viewpoint regarding offered products and services by firms in favor of developing innovative marketing strategies.

Among the mainstream SNSs, the popularity of twitter grows exponentially. Launched in 2006, Twitter has created a new social phenomenon and has attracted more than 288 million monthly active users as of 2014, posting 500 million tweets per day [Static brain (2013)]. Based on the twitter platform, a wide variety of research issues in mining twitter data has been investigated. Studies have explored extracting public sentiment [Pang and Lee (2008); Go et al. (2009); Das and Chen (2007); Pang et al. (2002)] in a variety of cases, but in this study we set out to look deeper by adding another important dimension, which is the content generator's impact on grabbing public (users) attention. This is due to the fact that some criteria's play an important role for a tweet being read and distributed. The number of retweets and followers of a profile are some of these essential factors. In this study, we aim not only to investigate the overall sentiment polarity of the public regarding a specific product category, but also to look deep into the top content which makes the company's image. For the purpose of content evaluation a survey was designed and crowd intelligence evaluation was collected regarding the top content related to a company's product during the period of investigation.

The paper is organized as follows: We start by discussing the emergence of Social Network Services (SNSs) and thus the importance of being innovative in SNSs from a firm's perspective. Advancements in computer sciences and text analysis techniques provide an opportunity to extract meaningful assumptions from content in SNSs. We then elaborate a case study of the top US smartphone manufacturers in Section 2 and construct the research questions. In Section 3, the attempt is to retrieve data from Twitter, and the searching queries for this purpose will be designed accordingly. More specifically, the top tweets were detected based on their generated impact, and the types of profile were categorized based on our five twitter profile types defined. Later on, sentiment analysis will be utilized to understand the polarity of tweets (positive or negative). In addition, a survey was designed in order to perform crowd intelligence evaluation regarding the

content of tweets. Finally, multiple variables were constructed and a correlation analysis was executed in order to explore the possible relationships. As a result, we shall develop a number of propositions on a firm's social media marketing strategies.

2. Literature Review

From the perspective of companies and businesses, being innovative in advertisement is a key challenge for firms to distinguish themselves and raise public awareness. With the diversification of media channels, the necessity of understanding each media channel, its characteristics, and its effect need to be considered. The development of Social Networking Services (SNSs) is one of the key phrases of the next generation internet. Web 2.0 provides a platform where new SNSs have emerged to facilitate the building of virtual relationships online. The broadening of the usage of social media in everyday life eventually made the social media an important channel for marketing. SNSs have become a rich source of User-Generated Content (UGC), and they are attracting a growing interest in various research domains, including sentiment analysis [Pang and Lee (2008)], scientometrics [Eysenbach (2011)], online user behavior [Fu and Shen (2014)], and user profile detection [Ikeda et al. (2013); Galán-García et al. (2014)].

Social network services connect businesses to end-consumers effectively at a low cost [He et al. (2013)], influence customer awareness and behavior, and bring together the likeminded [Hutton (1998)], which are the advantages that have made SNSs the center of attention in different industries. The unparalleled efficiency of SNSs compared to traditional communication channels prompted a statement from industry leaders encouraging companies to manage online environments by participating in Twitter and Facebook, among others [Kaplan and Haenlein (2010)].

While there is a surge in the amount of content being generated in SNSs, no clear indication of the quality of these types of content exists. Assessing the quality of the content is rather subjective and cannot be captured by current technologies. The challenge still remains to provide innovative and eye-catching content. Once in a while a video in YouTube gets viral which reminds us of the rapid changing of human taste for consuming content in social media [Zarella (2013)]. This will also alert companies about the fast evolution of this era and the need to adapt and to be innovative regarding their content and distribution channels.

To the extent of our knowledge, most studies have looked at the overall sentiment of the public in regard to a company's brand presence, without incorporating the source of the content and its embedded intention. The quality of content in social media is highly important, as intelligent individual cherry-pick content for further attention. Recent research has also recognized the importance of the issue, but the mechanisms for the filtering process are still biased and unknown. This will accentuate the importance of high quality content and, relevant to that, the content producer and the intention of sharing will play a remarkable role.

With the advancement of analyzing human language data and the tools widely available for this purpose, the opportunity is available to see how social media content is attached to real world events. Also, with the emergence of crowd intelligence services, which provide the human labor, we had the advantage and firsthand experience of outsourcing a load of

manual work. Due to complexity in profile detection, no absolute solution for profile categorization exists. Practices for detecting spam profiles [Thomas et al. (2011); Uddin et al. (2014)] are available, but the intention was to identify the profiles within the specific category of the concern of our study. This process requires additional attention and sources rather than twitter information itself.

In order to elaborate the importance of social media content and user perceptions of companies we conducted a case study. The focus of the case study was on smartphone manufactures in the United States in order to avoid geographical and cultural sampling errors. In the US, 169 million people own smartphones (70 % mobile market penetration). According to ComScore§ [Comscore (2014)], during the six months ending in July, Apple ranked as the top Original Equipment Manufacturer (OEM) with 41.9 % of US smartphone subscribers. Samsung ranked second with a 27.7 % market share, followed by LG with 6.6 %, Motorola with 6.1 %, and HTC with 5.1 %. Table 1 is showing the percentage share of the smartphone market in the US.

Table 1. Share (%) of smartphone OEMs.

	Feb-14	May-14	Jul-14	Average
Apple	41.3	41.9	42.4	41.9
Samsung	27.0	27.8	28.4	27.7
LG	6.8	6.5	6.4	6.6
Motorola	6.3	6.3	5.7	6.1
HTC	5.4	5.1	4.7	5.1

This study, examined the content exists on twitter for the five highest market shareholders in the smartphone industry for a particular time period. Natural Language Processing (NLP) techniques supplemented with sentiment analysis were applied to analyze unstructured text content on twitter. In addition, crowd intelligence was used to evaluate the quality of content perceived regarding the most impactful tweet's content relevant for each company's product. Furthermore, an inspection has been performed on the characteristics of the profile which generated the top tweets and their relative influence on explaining public sentiment polarity. Specifically, the study attempts to answer the following questions.

- How the content quality perceived from user perspective in SNSs varies among different type of content generators?
- How does the public sentiment reflect upon content's sentiment polarity within different category of content providers in SNSs?

The potential contribution of the paper can be linked to technology management tools as they play an important role in helping to make decisions in complex and dynamic environments [Phaal et al. (2006); Keltsch et al. (2011)]. Our study could provide an additional configuration for the technology management toolkit to understand the brand presence of firms in social networking services.

§ ComScore is an American internet analytics company providing marketing data and analytics to many of the world's largest enterprises, agencies, and publishers.

3. Data collection and methodology

3.1. *Social media data*

Twitter platform was selected for retrieving the company's product related textual data. A social networking and microblogging service, Twitter allows users to post real-time messages, tweets, restricted to 140 characters in length. The users of Twitter use the "@" symbol to refer to other users, which automatically alerts them. Hashtags "#" are used to mark topics primarily in order to increase the visibility of the tweets. Users can follow other users and/or be followed by other users (followers). Once a user follows another user, the generated tweets from the follower's user account will be automatically linked and shown on the user's home page. The unique characteristics and features of Twitter as a microblogging service are illustrated in Figure 1.

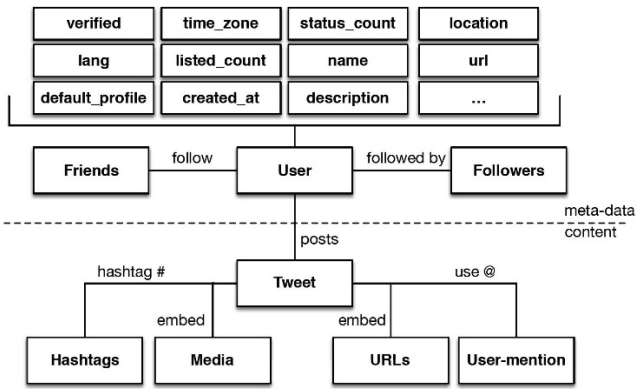


Figure 1. Twitter infographic diagram (adapted from Uddin et al. 2014).

With respect to twitter characteristics, the searching queries have been constructed in a way that captures the most relevant content regarding the specific product category of the studied companies. Desired products for each company has been selected from Phonearena.com (Premium website for mobile phone information and specifications). Study period has been set for 1 year, from January 2014 till end of December 2014. Geographical location has been fixed for United States and English language has been set. The constructed searching queries can be find in Appendix A.

At this stage we used the twitter search web interface** and broke the queries into various time intervals in order to allow an easier/smoother capturing process of past tweets. This consists of 12 iterative process (for 12 months) of running queries for each respected case company. The responding matches loaded into a modified light performance browser for better management of the system memory. In the next step, the loaded content with the same format as it displayed in the browser, transferred to an application that can show the same quality content as it's an important step for cleaning and categorizing data. In order to extract the required information easier with greater precision, a rich text editor application "MS Word" has been used. Tweets identical features such as username, time, and location plus textual formatting features such as text color and text size have been

** https://twitter.com/search-advanced

benefited for separating different data types automatically. For better data management and analysis, a tabular program had been used, in a way to operate on data represented as cells of an array, organized in rows and columns, widely known as the spreadsheet. "MS Excel" as a spreadsheet program had been used for this purpose. With some minor modifications for transferring the textual data into a sensible format.

The total amount of tweets retrieved for the targeted companies over the period was about 292,000 tweets. Apple smartphones related tweets consist 60% of all retrieved tweets, following that Samsung with 28%, HTC with 6%, Motorola with 4% and LG with 2%.

In the next section we will encounter the number of the followers of the twitter profiles, generating the tweets, in order to sort the tweets accordingly from the highest to the lowest. This is a promising way to identify the tweets with a high impact as the profile has a higher number of followers. Furthermore we shall detect different profile type of tweets.

3.2. *Types of Twitter Profiles*

The main motivation for twitter profile investigation mainly initiated from the fact that humans as intelligent individuals will impose complex factors where it comes to consume and disseminate information in SNSs. Moreover, the quality of content highly affiliates with the content generator characteristics and intention as different purposes create different needs; therefore each profile type serves a different purpose. Twitter profile categorization highly depends on the intention of the study. We therefore chose to construct five different types of twitter profile with below descriptions.

I. Personal profiles: These accounts contain personal content and have no tie or mention of corporate or brand information. They are created by individuals who do not want to be associated with their employer. Technically, the accounts have been created for the purpose of acquiring news, learning, fun, etc. Generally, these individuals show a low to mild behavior in their social interaction.

II. Professional profiles: These are personal users who communicate with professional intent on Twitter. They share useful information about specific topics and are involved in healthy discussion related to their area of interest and expertise. Professional users tend to be highly interactive: they follow many and are also followed by many.

III. Corporate and business profiles: These users are different from personal and professional users in that they follow a marketing and business agenda on Twitter. Their profile description strongly depicts their motive, and similar behavior can be observed in their tweeting behavior. Frequent tweeting and less interaction are the two key factors that distinguish business users from both personal and professional users. The type of content might be primarily corporate content. This account can be managed by a team, often sporting the proper brand name of a company, and provide corporate news, deals, and support.

IV. Feed/news profiles: These profile types represent automated services that post tweets with information taken from news websites such as CNN and BBC or from

different RSS feeds. The intentions of these tweets are news broadcasting, and they cover various areas.

V. Viral/marketing services and spam profiles: These are mostly automated computer programs (bots), run behind the profile. The rate of this profile appearing and disappearing is rather high, as Twitter can detect the patterns and block these profiles. They provide a high portion of twitter content but at the same time are not easy to see because their content might not get spread compared to other types.

Having extracted the twitter username from the data extraction task, we polled twitter's public timeline through the twitter API and retrieved user profile info such as the number of followers (followers_count). For each company we sorted the tweets according to the followers count of the corresponding account. Therefore, the tweet generated by the account with higher number of followers, was located at the top. Then we looked at the top 10 twitter accounts for each month for every company and classified these user profiles manually within our five defined categories. With manual annotation, the 600 shortlisted twitter accounts were classified within the five categories defined in section 3.2. The next section sets out to analyze tweet sentiment.

3.3. *Sentiment analysis*

Finding the polarity of public opinion from the retrieved tweets is facilitated by an automated Sentiment Analysis (SA) technique. The purpose of the implementation is to be able to automatically classify a tweet sentiment as a positive or negative for a big set of data. The implementation of sentiment analysis of tweets using Python and the Natural Language Toolkit (NLTK)[††] was adopted from Luce [(2012)]. To understand the perceived sentiment of tweet content, we needed a corpus of tweet ratings. The computer, however, needs a training corpus or documents with information to learn from, and increasing the volume of documents will leads to better results. In sentiment analysis, the training corpus always involves example documents annotated manually into categories. Having learned from example, the computer can apply the acquired knowledge to new documents (a hold-out corpus or training set) and classify them into sentiment categories. In order to construct a training set, a multi domain dataset of Blitzer et al. [(2007)] had been utilized. The database is a product reviews from six Amazon product domains (book, dvd, electronics, kitchen, music, video) which has been manually annotated and classified as either positive or negative. Instruction has been followed from Luce [(2012)] in order to create a classifier by extraction of relevant word features and labeling. The constructed variable 'training set' includes the labeled feature sets which contain a list of tuples which each tuple containing the feature dictionary and the sentiment string embedded to it. Now, we have our training set which can train our classifier for predicting the opinion polarity of unseen tweets. Figure 2 is an illustration of the steps for conducting the sentiment analysis.

[††] http://www.nltk.org/

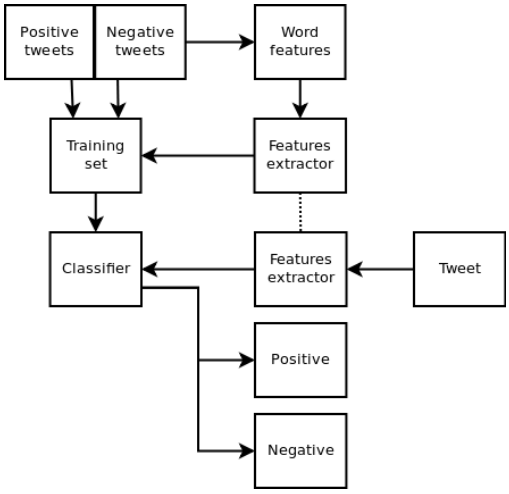


Figure 2 Sentiment analysis steps (adapted from Luce 2012).

The classification methods for performing sentiment analysis are many (e.g. Naive Bayes, Maximum Entropy, SVM). Naive Bayes is the algorithm which was used in our research in order to classify documents in categories of positive and negative sentiment. Bayesian classifiers are based on the Bayes rule which sees conditional probabilities so that the condition can be conveniently flipped around. A conditional probability is usually written P(X | Y) meaning that event X will occur, given evidence Y. This probability can be determined according to the Bayes rule when all we have is the probability of the opposite result and of the two components individually: P(X | Y) = P(X)P(Y | X) / P(Y). This facilitates estimating the probability of something based on examples of it occurring [Lewis (1998)].

Naive Bayes is a simple model which works well on text categorization [Weikum (2002)]. Even though text categorization based on the Naive Bayes model is simple and its conditional independence assumption is not valid in reality, it still performs surprisingly well [Lewis (1998)]. Domingos and Pazzani [(1996)] even showed that it is ideal for certain problem classes with highly dependent features. We tuned the Naive Bayes classifier in way that it uses the prior probability of each label which is frequency of each label in the training set, and the contribution from each feature.

Here, we are estimating the probability of a document being positive or negative, given its contents. This is convenient, because our data set has yielded examples of positive and negative opinions. Thus, the initial formula is

$$P(c|t) = \frac{P\ (c)P(t|c)}{P(t)} \tag{1}$$

where c is a specific class and t is text we want to classify. $P(c)$ and $P(t)$ are the prior probabilities of this class and this text, and $P(t \mid c)$ is the probability that the text appears given this class. Here, the value of class c might be positive or negative, and t is just a sentence.

The goal is choosing the value of *c* to maximize *P(c | t)*: Considering "*w*" as word, where *P(wi | c)* is the probability of the *i*th feature in text *t* appears given the class *c*. We need to train parameters *P(c)* and *P(wi | c)*. The Naive Bayes model produces these parameters effortlessly, since they are just the maximum likelihood estimations (MLEs) of each. When making a prediction about a new sentence t, we calculate the log likelihood P(c) +∑ilogP(wi|c) of different classes, and take the class with the highest log likelihood as prediction. In practice, this needs smoothing to avoid zero probabilities, which occur when there is an unseen word when making prediction [Narayanan et al. (2013)].

Let's go through an example of a text 'My phone is not good'. The first step will discard any feature names that are not known by classifier if exists. Next step is to find the log probability for each label. The probability of each label ('positive' and 'negative') is by default 0.5. Due to our implemented training set the word 'good' weights on the positive side but the word 'not' is part of more negative text in our training set so the output from the classifier is 'negative'. With the same logic we happen to see the following text: 'This phone is not bad' would return 'negative' even if it is 'positive'. Again, a large and well-chosen sample will help with the accuracy of the classifier.

In order to estimate the accuracy of the sentiment classifier on the training set, an accuracy measure has been calculated accordingly. As the trained model will be used to predict the polarity of an unseen tweet, knowing the accuracy ratio is important. In pattern recognition and information retrieval with binary classification, the accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined. Accuracy is measured by the area under the ROC (Receiver Operator Characteristic) curve which an area of 1 represents a perfect test; an area of 0.5 represents a worthless test. Figure 3 is an illustration of ROC curve which corresponds to our classifier model.

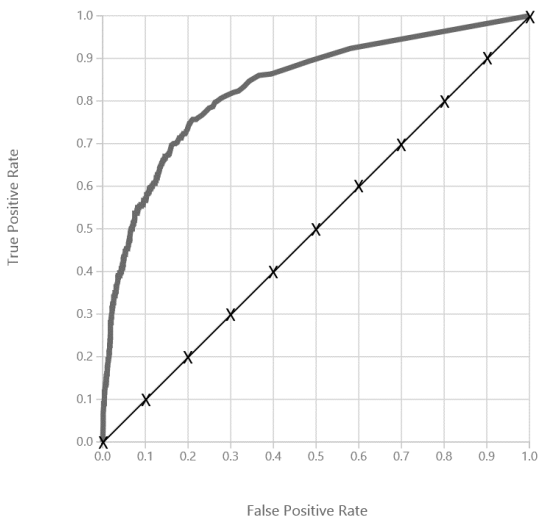


Figure 3. ROC curve.

Accuracy ratio has been used as metric to evaluate the usefulness of a model:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (2)$$

where

TP = Number of True Positive,
FP = Number of False Positive
TN = Number of True Negative,
FN = Number of False Negative

Our results from the acquired Naive Bayes classification algorithm indicates 76% of accuracy measure which is the area covered under the ROC curve. The acceptable accuracy rate suggests that using the Naive Bayes classifier for detecting tweet sentiment could be used in practice.

After the sentiment detection, a high share of neutral tweets is understandable and has been addressed by other researchers following the same practice [Kwak et al. (2010); Yu et al. (2013)]. We also ignored the neutral sentiment result and considered using the positive and negative detected results for the same reasons: subjective expressions always imply either positive or negative feelings. Moreover, from the methodological perspective (machine learning) the mature sentiment repository is not yet available to efficiently and accurately identify neutral sentiment. The detected portion of positive and negative sentiment for each company in each month is presented in Appendix B.

The result in this section helped us to understand the overall sentiment polarity carried by content on twitter regarding the company's product. In the next section the attempt is to evaluate the content on twitter from the human perspective. The calculated scores for the tweets sentiment transferred to spreadsheet for further analysis.

3.4. *Crowd intelligence content evaluation*

This subsection tries to investigate which type of tweets gets higher evaluation from individuals based on their producers. Within our initial shortlisting of twitter profiles with the highest number of followers, we tend to focus on contents of the top tweets from these profiles accordingly.

Due to the development of internet-based tools, we outsourced work to individuals in order to construct the survey. Technically we applied the concept of c*rowdsourcing* which has been defined as a practice of getting ideas and services from a large group of people willing to contribute their input. This type of outsourcing can be applied to various activities and most notably to tedious tasks [Safire (2009)].

In order to capture the level of the user's interest for these top tweets, a survey was designed, asking the likelihood of users reading the represented content. Mechanical Turk (MTurk), a service from Amazon, was used to perform the content evaluation task. A crowdsourcing internet marketplace, MTurk allows individuals and businesses (known as requesters) to coordinate the input of individuals in performing tasks that cannot be computerized. Tweets are notoriously hard to classify given all the abbreviations, sarcasm, and hash tags; therefore individuals may give the best results. Participants were asked to

evaluate their interest of reading a content on a Likert scale (1=not at all interested to 5=very interested). 300 people get involved with the questionnaire and evaluated the tweet's content quality. Each individual participant was provided with a randomly selected question (out of 120 variation of questions) to do the evaluation task. Appendix D is showing one sample of such questions. The detailed crowd content evaluation score can be seen in Appendix C whereas the distribution of the results can be seen in Appendix E which will be discussed further in the next section.

4. Findings

For answering the research questions, the calculated indices were considered, and the following variables were constructed in order to investigate the relationships between them. Table 2 shows the constructed variables with their description.

Table 2. Cconstructed variables with their description.

Variable	Description
PERSON_USER_COUNT it	Total tweets posted by personal profiles for company i in month t
PERSON_USER_POS it	The number of positive sentiment tweets by personal profiles for company i in month t
PERSON_USER_NEG it	The number of negative sentiment tweets by personal profiles for company i in month t
PROF_USER_COUNT it	Total tweets posted by professional profiles for company i in month t
PROF_USER_POS it	The number of positive sentiment tweets by professional profiles for company i in month t
PROF_USER_NEG it	The number of negative tweets by professional profiles for company i in month t
COAP_USER_COUNT it	Total tweets posted by corporate and business profiles for company i in month t
COAP_USER_POS it	The number of positive sentiment tweets from corporate and business profiles for company i in month t
COAP_USER_NEG it	The number of negative sentiment tweets from corporate and business profiles for company i in month t
NEWS_USER_COUNT it	Total tweets posted by feed/news profiles for company i in month t
NEWS_USER_POS it	The number of positive sentiment tweets from feed/news profiles for company i in month t
NEWS_USER_NEG it	The number of negative sentiment tweets from feed/news profiles for company i in month t
VIRAL_USER_COUNT it	Total tweets posted by viral/marketing services and spam profiles for company i in month t
VIRAL_USER_POS it	The number of positive sentiment tweets from viral/marketing services and spam profiles for company i in month t
VIRAL_USER_NEG it	The number of negative sentiment tweets from viral/marketing services and spam profiles for company i in month t
OVERALLCAT_NEG_SENT it	The number of negative sentiment detected in top tweets in 5 categories for company i in month t
OVERALLCAT_POS_SENT it	The number of positive sentiment detected in top tweets in 5 categories for company i in month t

Refereeing to the profile detection and categorization task which explained in section 3.2, figure 4 is a representation of the 600 twitter profiles which has been detected and categorized manually. As it can be seen from the figure, the distribution of the detected profiles for each company case is very different. Companies such as Apple, Samsung and LG has bigger portion of professional content producers whereas for HTC and Motorola the dominance category is News and Corporate account types. Relatively, the proportional

figure for viral type of account is very miner in all companies and that mainly is due to the mechanism that the tweets were shortlisted for profile categorization task. The shortlisting mechanism was based on the number of followers of the twitter profile which eventually is a proxy for interesting content that will left out the spam type of a content.

The right side of the figure 4 is an illustration of crowd evaluation survey regarding the quality of a tweet content. The fact that can be depicted from the crowd intelligence evaluation survey, result is that the participants found the contents relatively interesting to read as the evaluation results are skewed towards the right. This can be understandable as long as the content's material for the survey where shortlisted based on the impact of the content producers which is a proxy for a content being interesting. However, what can be observed from the results is that *professional user* content is more interesting and in the second place the pure *personal* content is interesting for participants to read. The other point worth to be noted is the corporate content that have the lower valuation score from participants.

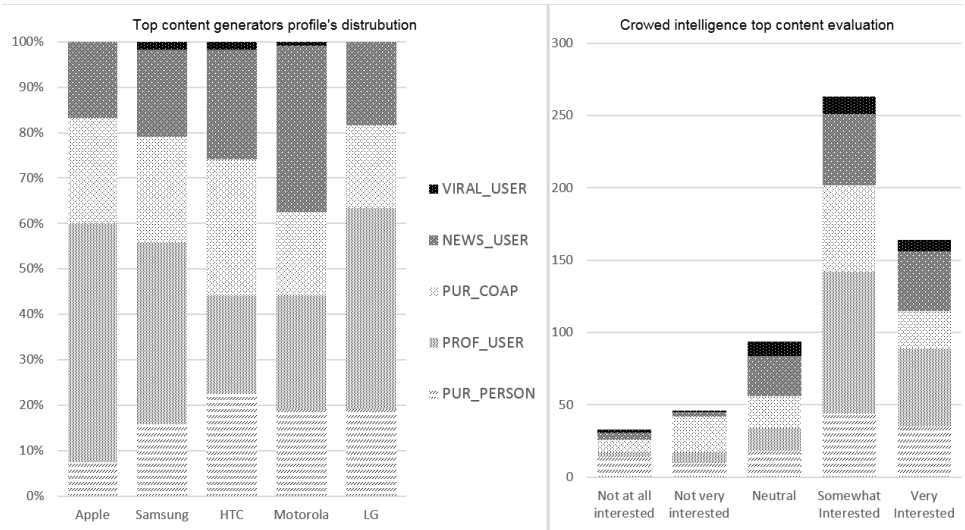


Figure 4. Profile distribution and content evaluation.

Regarding the second research question, a test has been designed to investigate the influence of top contents (tweets) to the public sentiment. Therefore, for each individual case company, a Pearson's product-moment correlation was ran to assess the relationship between public sentiment and overall categorical sentiment detected for each company. Table 3, shows the result of the Pearson's correlation analysis between public sentiment as dependent variable and the overall categorical sentiment calculated for each individual company as independent variable. The test had been handled in separate for positive and negative detected tweets. The correlation matrix has been simplified and the results are represented in table 3.

Table 3. Pearson's correlation estimation of public sentiment and categorical sentiment.

Variable	Pearson Correlation				
	Apple	Samsung	HTC	Motorola	LG
Overall public positive sentiment as dependent variable					
OVERALLCAT_POS_SENT it	.233	.037	.068	.459	-.021
	(.465)	(.910)	(.833)	(.134)	(.949)
Overall public negative sentiment as dependent variable					
OVERALLCAT_NEG_SENT it	.785**	.583*	.552	.867**	.860**
	(.002)	(.047)	(0.63)	(.000)	(.000)

**. Statistically significant at p < .01 level
*. Statistically significant at p < .05 level

There is a statistically significant relationship for 4 case companies. The results impose the different carrying capacity of negative and positive high impactful content (tweets).In Table 4 we extend the analysis in order to estimate how the overall public sentiment result explains the relationship between the sentiments within the contents of each particular twitter profile category we detected earlier.

Table 4. Correlation coefficient estimation for overall twitter sentiment results with twitter top content sentiment

Variable	Pearson Correlation				
	Apple	Samsung	HTC	Motorola	LG
Overall positive sentiment as dependent variable					
PERSON_USER_POS *it*	-.550	-.168	.069	.195	-.192
	(.064)	(.603)	(.831)	(.544)	(.551)
PROF_USER_POS *it*	.226	.117	-.315	-.033	-.298
	(.480)	(.718)	(.318)	(918)	(.347)
COAP_USER_POS *it*	.088	-.260	.511	-.166	.321
	(.785)	(415)	(.090)	(.605)	(.309)
NEWS_USER_POS *it*	.480	.223	-.207	.393	.103
	(.114)	(486)	(.519)	(.207)	(.750)
VIRAL_USER_POS *it*	--	--	-.578*	--	--
			(.049)		
Overall negative sentiment as dependent variable					
PERSON_USER_NEG *it*	--	.747**	-.056	.660*	.549
		(.005)	(.864)	(.020)	(.065)
PROF_USER_NEG *it*	.932**	-.381	.365	.097	.794**
	(.000)	(.221)	(.244)	(.765)	(.002)
COAP_USER_NEG *it*	-.464	.226	--	--	--
	(.129)	(.480)			
NEWS_USER_NEG *it*	-.267	-.457	.633*	.048	-.210
	(.401)	(.135)	(.027)	(.881)	(.513)
VIRAL_USER_NEG *it*	--	--	--	--	--

**. Statistically significant at p < .01 level
*. Statistically significant at p < .05 level

Results in table 3 revealed significant relationship between negative sentiment from public and negative categorical sentiment captured in top tweets. Hereby, we are expecting to see how the strong relationships appear in categories of profile within top tweets. As it can be seen from table 4, in four out of five case companies, personal and professional users' tweets carrying negative sentiment had a positive relationship with the overall negative sentiment captured for each company. This result burdens the fact of the weight and value of negative massages compared to the positive and neutral ones. The moderate effect of negative news and its role in financial markets has also been captured and explained in earlier studies [Xin Xu and Zhang (2013)].

5. Discussion and Conclusion

Crowd evaluation of company's related content in Social Network Services (SNSs) revealed the importance of its quality rather than its quantity. This magnifies the importance of content representing companies in SNSs and emphasizes on the innovative ways of producing and delivering it.

In this study we utilized sentiment analysis techniques in order to understand the overall polarity of public sentiment regarding specific companies' products. At the same time, a crowd intelligence evaluation of top tweets for each company was also made as we recognized the importance of the content producer and his or her intent to share it in social media. Adding to the literature on innovation in social media marketing, our findings introduce the content producer's impact on shaping a company's image through public opinion. While being innovative in social media is rather vague and nearly impossible to capture, we illustrate how the users of social media prefer to look at more human and generic content rather than pure advertisements and corporate announcements.

Specifically, we showed that overall sentiment perceived from content in Social Network Services (SNSs) does not necessarily relate to what normal users might perceive of a brand. When viewing content providers categorically in social media (Twitter), professional users who were not affiliated with any company got higher scores in crowd intelligence evaluation for getting their content read.

Next we investigated, how public sentiment sensitivity is triggered by most interested content in SNSs. The results had been indicated that negative sentiment polarity coming from the top tweets has more explanatory power to explain the public sentiment. The same effect has not been shown in categorical positive sentiment. In a more detailed view, negative sentiment detected in personal and professional content showed a positive relationship with overall negative sentiment. This again amplifies the importance of this humanistic type of content producers in social media.

The impact of negative tweet's sentiment indicates the importance of top content in terms of representing a company's image in social media. Admittedly, the overall sentiment result may not be a sufficient indicator for understanding the company's image among the public. Due to the fact that all the content (tweets) is not subjected to equal audiences, the content generator plays a significant role as a distribution channel which finally contributes to forming the company's image in SNSs.

In summary, our results suggest that sentiment analysis can contribute to our ability to understand the user's perceived understanding of a company's specific products. Importantly, a detailed view of content generators and the level of their influence in social media would highly add to this understanding. While information exchange mechanism is different when it comes to other type of SNSs platforms, a future study could deeper our understanding of company's image in other type of SNSs channels.

6. Managerial Implications

The present work provides theoretical contributions with a relevant impact on a variety of contiguous fields. It encompasses the elaboration of innovative marketing methods with social network services and can help corporations to adjust their advertising strategies. Marketing in social media has complex logistics. Currently, it is the matter of good content,

good timing, and the right distribution channel. As the public showed a high tendency to look for more humanistic profile types such as personal and professional accounts, a company with different twitter accounts makes a brand more preferable to individuals and can help to bond a connection. Meanwhile, this strategy of multiple accounts provides the customers with a great service portal. Their success is evident as more and more brands are adopting this approach.

Playing on the strengths of each, the different types of profiles are useful for a branding strategy. Internal coordination with a process and policy, however, will also help to provide a common and high-quality experience to customers. It is vital to understand which type is right for the company's social media, where the culture and goals of the organization set the course. Companies can have several twitter accounts to serve different purposes. One account can be used to share factual information, whereas another can be in a supportive role and the other one can meant to help employees who have little connection to the product or customers. To ensure the quality and effectiveness of these accounts, however, it is important to establish beforehand whether they have an audience and that enough effort is put to keeping them up-to-date with fast response rates.

7. Acknowledgements

We thank the senior editor, and anonymous reviews for their constructive feedbacks which enable us to improve the paper significantly. We thank Niklas Ahlgren and Mahmut Yasar for assistance with statistical test selection and further comments that greatly improved the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed persons.

References

Blitzer, J. et al.,(2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. *Annual Meeting-Association for Computational Linguistics*, **45**, p.440.

Comscore,(2014). comScore Reports May 2014 U.S. Smartphone Subscriber Market Share. [Accessed April 12, 2015]

Das, S.R. & Chen, M.Y.,(2007). Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science*, **53**(9), pp.1375–1388.

Domingos, P. & Pazzani, M.,(1996). Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier. *Machine Learning*, pp.105–112.

Elkin, N.,(2008). How America searches: health and wellness. *Opinion Research Corporation.*

Eysenbach, G.,(2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of medical Internet research*, **13**.

Fu, X. & Shen, Y.,(2014). Study of collective user behaviour in Twitter: a fuzzy approach. *Neural Computing and Applications*, **25**, pp.1603–1614.

Galán-García, P. et al.,(2014). Supervised Machine Learning for the Detection of Troll Profiles in Twitter Social Network : Application to a Real Case of Cyberbullying. *International Joint Conference SOCO'13-CISIS'13-ICEUTE'**13***, pp.419–428.

Go, A., Bhayani, R. & Huang, L.,(2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*.

He, W., Zha, S. & Li, L.,(2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, **33**(3), pp.464–472.

Holtzblatt, L.J.,(2011). Measuring the Effectiveness of Social Media on an Innovation Process. *Idea*, pp.697–712.

Hutton, G.,(1998). Net gain: Expanding markets through virtual communities. *Long Range Planning*, **31**, pp.328–329.

Ikeda, K. et al.,(2013). Twitter user profiling based on text and community mining for market analysis. *Knowledge-Based Systems*, **51**, pp.35–47.

Kaplan, A.M. & Haenlein, M.,(2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, **53**, pp.59–68.

Keltsch, J.N., Probert, D. & Phaal, R.,(2011). A process for configuring technology management tools. *International Journal of Technology Intelligence and Planning*, **7**, p.181.

Kwak, H. et al.,(2010). What is Twitter, a social network or a news media? *19th international conference on World wide web*, pp.591–600.

Lewis, D.D.,(1998). Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval. *Machine Learning: ECML-**98***, pp.4–15.

Luce, L.,(2012). Twitter sentiment analysis using Python and NLTK. Available at http://www.laurentluce.com/posts/twitter-sentiment-analysis-using-python-and-nltk/ [Accessed March 10, 2015]

Nair, M.,(2011). Understanding and Measuring the value of social media. *Wiley periodicals*, pp.45–51.

Narayanan, V., Arora, I. & Bhatia, A.,(2013). Fast and accurate sentiment classification using an enhanced Naive Bayes model . , pp.1–5.

Pang, B. & Lee, L.,(2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, **2**(1), pp.1–135.

Pang, B., Lee, L. & Vaithyanathan, S.,(2002). Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics.*

Phaal, R., Farrukh, C. & Probert, D.,(2006). Technology Management Tools: Generalization, Integration and Configuration. *International Journal of Innovation and Technology Management*, **03**(3), pp.321–339.

Safire, W.,(2009). crowdsourcing. *The new york times*. http://www.nytimes.com/2009/02/08/magazine/08wwln-safire-t.html4&ref=?_r=magazine& [Accessed April 12, 2015].

Static brain (2013). Twitter Statistics. Twitter statistics. Available at http://www.statisticbrain.com/twitter-statistics/ [Accessed April 5, 2015].

Stelzner, M.A.,(2013). How Marketers Are Using Social Media to Grow Their Businesses. *Social Media Marketing Industry Report*

Thomas, K. et al.,(2011). Suspended accounts in retrospect: an analysis of twitter spam. *Proceedings of the 2011 ACM Internet Measurement*, pp.243–258.

Uddin, M.M., Imran, M. & Sajjad, H.,(2014). Understanding Types of Users on Twitter. *In Proc. of the SocialCom-Stanford conference*, p.6.

Weikum, G.,(2002). Foundations of statistical natural language processing. *ACM SIGMOD Record*, 31, p.37.

Xin Xu, S. & Zhang, X. (Michael),(2013). Impact of Wikipedia on Market Information Environment: Evidence on Management Disclosure and Investor Reaction. *MIS Quarterly*, **37**, pp.1043–A10.

Yu, Y., Duan, W. & Cao, Q.,(2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, **55**(4), pp.919–926.

Zarella, D.,(2013). *The Social Media Marketing book*, O'Reilly Media.

Appendix A

Table A.1. Twitter searching queries

Company	Product	Searching query
Apple	Iphone 5, 5s, 5c, 6, 6 plus	Results for Iphone 5s OR 5c OR 6 OR plus lang:en near:"United States" since:2014-1-1 until:2014-12-30
Samsung	Galaxy S series including II III 4, note	Results for samsung Galexy OR II OR III OR 4 OR note lang:en near:"United States" since:2014-1-1 until:2014-12-30
LG	G3, G2, G pro 2, G flex, optimus G pro	Result for LG OR G3 OR G2 OR G OR G pro 2 OR G flex OR optimus G pro lang:en near:"United States" since:2014-1-1 until:2014-12-30
Motorola	Droid series, MOTO	Result for Motorola OR Droid series OR MOTO lang:en near:"United States" since:2014-1-1 until:2014-12-30
HTC	ONE, DESIRE	Result for HTC OR one OR desire lang:en near:"United States" since:2014-1-1 until:2014-12-30

Appendix B

	Sentiment	Jan-14	Feb-14	Mar-14	Apr-14	May-14	Jun-14	Jul-14	Aug-14	Sep-14	Oct-14	Nov-14	Dec-14
Apple	*Negative*	37%	48%	54%	27%	32%	49%	58%	45%	25%	30%	48%	31%
	Positive	63%	52%	46%	73%	68%	51%	42%	55%	75%	70%	52%	69%
Samsung	*Negative*	52%	55%	30%	31%	27%	74%	25%	42%	47%	40%	40%	65%
	Positive	48%	45%	70%	69%	73%	26%	75%	58%	53%	60%	60%	35%
LG	*Negative*	42%	40%	48%	37%	35%	55%	26%	29%	25%	63%	18%	22%
	Positive	58%	60%	52%	63%	65%	45%	74%	71%	75%	37%	82%	78%
Motorola	*Negative*	70%	65%	63%	70%	82%	63%	65%	60%	76%	58%	57%	58%
	Positive	30%	35%	37%	30%	18%	37%	35%	40%	24%	42%	43%	42%
HTC	*Negative*	52%	50%	42%	56%	35%	22%	57%	63%	37%	42%	73%	35%
	Positive	48%	50%	58%	44%	65%	78%	43%	37%	63%	58%	27%	65%

Appendix C

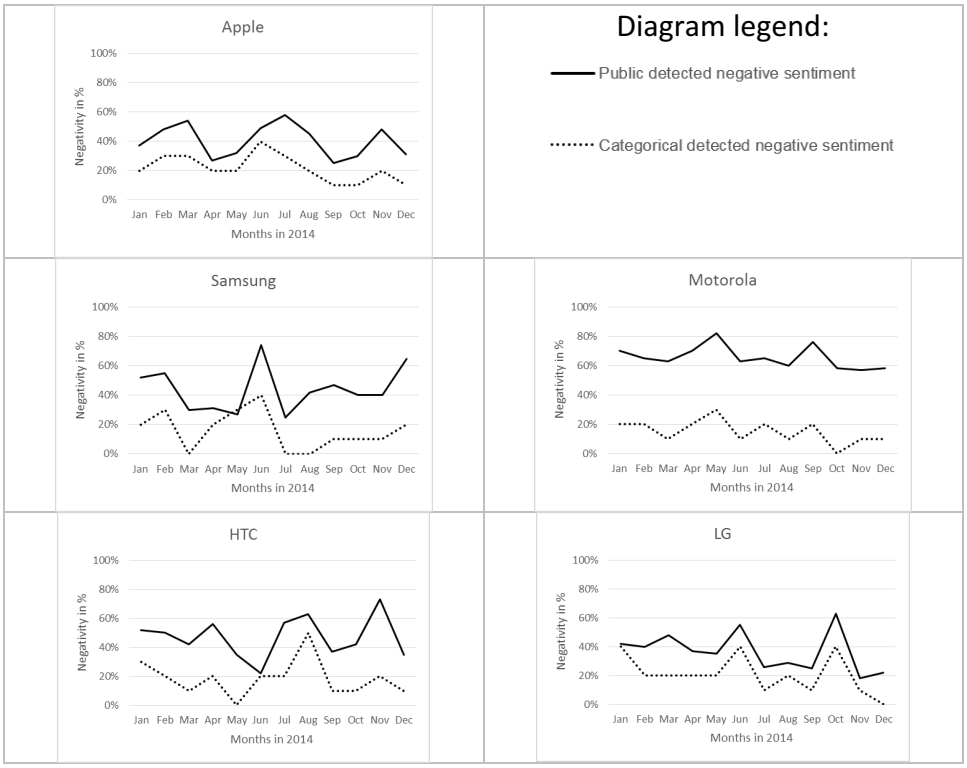
	Jan-14	Feb-14	Mar-14	Apr-14	May-14	Jun-14	Jul-14	Aug-14	Sep-14	Oct-14	Nov-14	Dec-14	Overall
Apple	0.60	0.80	0.73	0.60	0.33	0.67	0.47	0.87	0.67	0.80	0.73	0.60	0.66
Samsung	0.60	0.47	0.60	0.80	0.80	0.47	0.67	0.60	0.53	0.47	0.53	0.33	0.57
LG	0.53	0.47	0.33	0.60	0.67	0.73	0.73	0.87	0.47	0.60	0.33	0.40	0.56
Motorola	0.33	0.60	0.47	0.47	0.33	0.67	0.60	0.80	0.47	0.40	0.67	0.80	0.55
HTC	0.40	0.53	0.53	0.53	0.47	0.67	0.40	0.33	0.53	0.67	0.33	0.53	0.49

Appendix D

Please specify the level of your interest for reading the 5 Tweets option provided.

	Not at all interested	Not very interested	Neutral	Somewhat interested	Very interested
Option 1	○	○	○	○	○
Option 2	○	○	○	○	○
Option 3	○	○	○	○	○
Option 4	○	○	○	○	○
Option 5	○	○	○	○	○

Appendix E



Biography

Arash Hajikhani M.Sc. (Tech.), is a Researcher and a Doctoral Student in School of Business and Management at Lappeenranta University of Technology (LUT). He has been studying methods for evaluating and measuring various types of innovation. In his Master's thesis, he conducted a study on transforming unstructured data in social media into a validated indicator for measuring innovation. His current research interest is in analyzing different regional innovation ecosystems; finding their main entities/players and the interaction between them.

Jari Porras D.Sc. (Tech.), is Professor of Software Engineering (especially distributed systems) at the Lappeenranta University of Technology (LUT). Professor Porras received the D.Sc. (Tech.) degree from the Lappeenranta University of Technology, Finland in 1998 about modeling and simulation of communication networks in distributed computing environment. He has been working and publishing on various aspects of wireless networks and services, software and application development as well as educational development. In recent years, he has conducted research on green IT and sustainable ICT. He is interested in sustainable software innovations and is the person responsible in LUT for the Erasmus Mundus Perccom program.

Helinä Melkas D.Sc. (Tech.), is Professor of Industrial Management (especially service innovations) at Lappeenranta University of Technology (LUT). She has an interdisciplinary background with degrees of Master of Arts and Licentiate of Social Sciences. She has worked at LUT Lahti since 2007. In recent years, Helinä Melkas has conducted research on, inter alia, innovation management at policy and organizational levels, knowledge management and technology assessment. She has published tens of articles in international and national scientific journals and books.

Publication IV

Hajikhani, A.

Efficiency Assessment of the Social Capital Capacity on Entrepreneurial Activity: A Perspective Driven From Social Media

Reprinted with permission
2017 Portland International Conference on Management of Engineering and Technology (PICMET)
Portland, OR, 2017, pp. 1-10.
doi: 10.23919/PICMET.2017.8125317

Efficiency Assessment of the Social Capital Capacity on Entrepreneurial Activity: A Perspective Driven from Social Media

Arash Hajikhani
Lappeenranta University of Technology, Finland

***Abstract*— *Innovation is the main engine of a sustained economic growth that leads to a vibrant economy. Meanwhile, the large and diverse array of participants and resources are constructing an ecosystem where the synergies contribute to ongoing innovation and flourish in a modern economy. One major dimension of a healthy innovation ecosystem is illustrated in the social-economical aspects of society known as entrepreneurship. Our research concern is to capture the efficiency of social capital capacity towards entrepreneurship-oriented activities in society at large. The approach is to look at the perspective in the form of an input and output where inputs are the governmental efforts for educating human capital in the society and the output is the entrepreneurial-oriented activities and desires. Special focus was given to social network services in order to capture entrepreneurial activity as well as leverage established reports in the country level performance benchmarking practices regarding innovation and entrepreneurship. The level of analysis is country level by looking at a sample of European countries. In order to assess the efficiency of input efforts to be transformed into output, a non-parametric method dominant in operation research and economics known as Data Envelopment Analysis (DEA) has been utilized. Efficiency measures are calculated which generate a new scale and ranking accordingly, with an emphasis on efficiency rather than proficiency of social capital capacity.*

I. INTRODUCTION

Innovation and entrepreneurship concepts are highly intertwined and dependent on each other [1]. In a seminal work in the field of Business, Economics and Management, Drucker focused on two aspects of innovation: the process of innovation (i.e. how innovators search for opportunities and transform them into a new practice in the marketplace) and the practice of "entrepreneurship" (i.e. how an organization embeds the practice through institutional processes) [2].

When two disciplines are so closely aligned, as are innovation and entrepreneurship, in order to get actionable information, there is a need to look at entrepreneurship and innovation interchangeably. Consequently, various scholars have attempted to calculate, capture, compare and report measurements of innovation and entrepreneurship activities in the form of an index. The index compares a region or county for assessing and benchmarking innovation and entrepreneurship capacity, and is commonly used by corporate and government officials to compare countries. The reports and assessments are in the form of an annual ranking of countries by their capacity for, and success in, innovation or entrepreneurship. The reports comprise surveys and in-depth interviews to measure the partnerships with various organizations and institutions in order to calculate the measures. GII[1] (Global Innovation Index) and GEM[2] (Global Entrepreneurship Monitor) are some recognized reports on evaluating countries' innovation level and their entrepreneurial activity. The mentioned annual reports measure various aspects of an economy (country). Their measurement varies from economical aspects to social and cultural structures of the countries.

The holistic and systemic perspective of an economy's performance in innovation and entrepreneurship provided from the mentioned report, triggers the question of the most influential measures among the many introduced. The quality and education level of human capital is one of the

[1] www.gemconsortium.org/

[2] https://www.globalinnovationindex.org/

978-1-890843-36-6 ©2017 PICMET

crucial factors for a successful innovation and entrepreneurial ecosystem. According to recent research, start-ups which had higher education among their founders had higher survival rate [3]. The importance of human capital has been pointed out by reports like GII in the year 2014. Meanwhile researchers have discussed the phenomenon from other dimensions such as culture. Culture is to the organization what personality is to the individual: "a hidden, yet unifying theme that provides meaning, direction, and mobilization" [4]. In the literature on the Quadruple helix proposed by Carayannis and Campbell, the focus is on social capacity [5]. In the fourth helix, particular attention is on highlighting the importance of human capital and large social capital in fostering innovation. Richard Florida has recognized the importance of embedded capacity in human capital and has coined the term "creative class" as a key driving force for economic development of post-industrial cities in the United States. Florida's work proposes that a new or emergent class of knowledge workers, intellectuals and artists is an ascendant economic force, representing either 1.) a major shift away from traditional agriculture or industry-based economies, or 2.) a general restructuring into more complex economic hierarchies [6]. While it is important to encounter the social and human capital capacity in to the analysis, it's hard to capture the influence of society with regards to entrepreneurship and entrepreneurial activity. Therefore, one flaw within the current reporting is that the attempt to use new data sources such as social media outlet to capture such a dynamic construct hasn't been seriously considered.

With the emergence of social media and widespread availability of access to the internet which led to the development of new forms of communication, vast opportunities are forming for capturing mass social interactions on social media platforms. Statista, an online portal for market data, estimates there will be some 2.67 billion social media users around the globe in 2018, up from 1.91 billion in 2014 [7]. Another Web analytics company, Compete, reports that among the top 10 websites in 2010, social networking sites (SNS) accounted for some 75% of total page views in the US, up from 31% in 2001 and 40% in 2006 [8]. The increased worldwide usage of smartphones and mobile devices is leading to high user engagement rates with SNS platforms [9]. SNS employ mobile and web-based technologies to create highly interactive platforms, ultimately enabling individuals and communities to share, co-create and discuss. The importance of social media outlets as valuable data sources establishes by the feature of pulsing societal concerns, while enable the further study and investigation on interactions between individuals and communities on social media platform.

In this study, the focus will be first on ways to capture the capacity of society in an economy in terms of its impact on entrepreneurial oriented activities. Second, the investigation will be accompanied by integrating data points which identify the activities on social network services platforms related to entrepreneurship. The methodological approach in the study will implement a benchmarking model to evaluate the operational efficiency of the resources invested and the relative outcomes in the context of European countries.

To guide our study, we set the following research questions:

- How to operationalize social network services data to evaluate the social capital capacity of an economy regarding entrepreneurial-oriented activities?
- How to evaluate the efficiency of an economy based on its social capacity towards its entrepreneurial-oriented activities?

This paper is structured as follows: In the following section, the importance of social capital and its educational capacity will be discussed and the previous literature will be reviewed. Further investigation and experiments will be conducted on a social network service platform in order to capture entrepreneurial-oriented activity in a set of countries. Next, the Data Envelopment Analysis model will be constructed for benchmarking of the studied countries. The process will carefully disclose capturing of the social capital capacity in terms of education, entrepreneurial intention and activity of the studied countries. The calculated efficiency scores and the resulted rankings in quantifiable terms in an analytical frame of DEA analysis will provide a new perspective on economies' performance on utilizing their social capital resources toward entrepreneurial-oriented activity and innovation.

II. LITERATURE REVIEW

Entrepreneurship and innovation are recognized as critical factors for the wealth and

competitiveness of a nation. Entrepreneurship, or the act of entrepreneurs, are crucial in any innovation ecosystem. Innovation is an inherently human endeavor, and successful innovation happens when people with skills, experience, and capabilities come together to understand or predict, and then address, other people's challenges. Talent, like capital and technology, is a key success factor for innovation, and inspiring potential talent will drive innovation and growth. The systems approach has been used to describe the multifaceted nature of innovation at various levels - national, regional, technological, and sectorial. The systems approach recognizes the interaction among the many actors and other "determinants of innovation processes, that influence the development and diffusion of innovations" [10]. This approach leads to understanding the processes by which research capabilities build knowledge, then transfer the knowledge to support business development; these processes are often understood in the context of the Triple Helix of business, government and academic interaction [11]. Earlier discussion of the 'Quadruple Helix' included the additional perspective of the media-based and culture-based public. What results is an emerging fractal knowledge and innovation ecosystem configured for the knowledge economy and society.

The complexity of designing and studding innovation programs have been risen to a new level and the reason assumed to be the systematic approach toward studding such a complex phenomenon. The previous simplified linier models of explaining the innovation such as research and development (R&D) expenditure is no longer adequate [12]. Indicators previously adapted from the assessment of industrial R&D evaluation are no longer sufficient to display the influence of programs on various parts of the innovation system and a systematic perspective requires more sophisticated and comprehensive means of evaluating the effectiveness of an innovation program [13]. Meanwhile, a major factor for this change of innovation process is the quality of human capital linked to the innovation activities carried out in countries. Other factors, such as technology and capital, also influence the innovation process; however, these directly correlate with the human factor, and therefore nurturing human capital will result in increased capacity for technology and other kinds of capital, leading to a solid foundation for innovation. Entrepreneurship and the human element in innovation can lead to better results at all levels.

The definition of human capital in the Oxford Dictionary is the skills, knowledge, and experience possessed by an individual or population, viewed in terms of their value or cost to an organization or country [14]. In an organizational context, this capital is the constantly renewable source of creativity and innovativeness, and human capital refers to the collective value of the organization's intellectual capital (competencies, knowledge, and skills) in the form of its employees [15]. On the other hand, structural and organizational capital includes "all the non-human storehouses" of knowledge within a firm [16], ranging from information systems, databases, and intellectual properties to culture-carrying artefacts of organization [17]–[19]. From organizational perspective, strong human capital is needed to leverage the ability of the firm as a whole organization to perform certain activities to a superior level relative to competitors [20], and by investing in human capital and its proactiveness and innovativeness, the organization will benefit through increased performance. This corresponds to the views of Bateman and Crant [21] regarding the role of employees in proactive firm behavior.
From an economical perspective, Adam Smith in his seminal work "Wealth of Nations" recognized the notion of human capital as the forth type of capital, which defied it as the acquired and useful abilities of all the inhabitants or members of the society [22]. Adam Smith articulates "The acquisition of such talents, by the maintenance of the acquirer during his education, study, or apprenticeship, always costs a real expense, which is a capital fixed and realized, as it were, in his person. Those talents, as they make a part of his fortune, so do they likewise that of the society to which he belongs" [23]. Alternatively, human capital is a collection of all of the knowledge, talents, skills, abilities, experience, intelligence, training, judgment, and wisdom possessed individually and collectively by individuals in a population. It is an aggregate economic view of the human being who is acting within economies; this view is an attempt to capture the social, biological, cultural and psychological complexity as they interact in explicit and/or economic transactions [24].

From sociological perspective, recent writings depict the notion of social capital, which has

extended the concept from an individual asset to a feature of communities and even nations. Another similar concept of entrepreneurial capacity was discussed in literature from an economical viewpoint, indicating that the increase in the quality of social capital causes not only a surge in economic activity, but also redistribution in favor of future generations [25]. From organizational perspective, entrepreneurial capacity allows a firm to capitalize on a broad scope of fresh, alternative perspectives that may fundamentally challenge embedded assumptions and path-dependent cognitive schemas that a firm uses [26]. We conjectured that the motivating factor for such entrepreneurial capacity might be the desire to acquire human capital and its accumulation which thereby influences the social capital.

Understanding how social capital leads to human capital is essential. The acquisition of social capital requires deliberate investment of both economic and cultural resources. Moreover, social capital may have its greatest impact in the accumulation of human capital, including the skills and knowledge that allow individuals to perform economically valuable labor [27], [28]. Education is considered as the primary vehicle through which people acquire human capital and achieve upward mobility, but what is the equivalent for social capital?

Social capital development on the internet via social networking websites such as Twitter or Facebook tends to be bridging capital according to one study, though "virtual" social capital is a new area of research [29]. Another perspective holds that the rapid growth of social networking sites suggests that individuals are creating a virtual network consisting of both bonding and bridging social capital. Twitter is currently one of the most popular social networking sites and touts many advantages to its users including serving as a social lubricant for individuals who otherwise have difficulties forming and maintaining both strong and weak ties with others [30].

Jih-Hsuan et al. [31] offer a noteworthy application of the scale of social media by measuring international residents originating from locations outside of the United States. The study found that social media platforms like Facebook provide an opportunity building social capital by connecting with Americans before arriving and then maintaining old relationships from home upon arriving to the States. The ultimate outcome of the study indicates that social capital is measurable and is a concept that may be operationalized to understand strategies for coping with cross-cultural immersion through online engagement.

There is no widely held consensus on how to measure social capital, which has become a debate in itself; this lack of consensus is a barrier for understanding how social capital affects the creation of human capital. In the context of this experiment, social capital by the act of an individual's participation in new communication outlets like social media platforms regarding entrepreneurial-oriented discussions will be recognized and defined. Effective adoption and execution of activities in social media is likely to require for an investment in social capital regarding proactiveness in entrepreneurial-oriented manners in society.

The study will carefully select the required variables for conducting an efficiency comparison of the studied countries. One of the main benchmarking methods widely applied is Data Envelopment Analysis (DEA) which has been a popular research topic in the previous decade and will be leveraged in this study. Data Envelopment Analysis (DEA) traces its origin back more than two decades to Charnes, Cooper, and Rhodes' 1987 paper [32]. The DEA model in its original form represented the performance or efficiency of the decision-making unit as the ratio of weighted outputs to weighted inputs, and has seen rapid expansion in recent years.

III. ENTREPRENEURIAL ACTIVITY IN SOCIAL NETWORK SERVICES

The popularity of social networking services consistently rises and new uses for the technology are frequently being observed. Web 2.0 tools such as Twitter provide the opportunity to gather novel metrics, accessible through application programming interfaces (APIs), which allows for a steady stream of data that is easily accessible [33], [34].

Twitter has become one of the most popular micro-blogging services on the Internet. Communication on Twitter occurs via small text-based messages of up to 140 characters. There is a small but quickly growing body of literature

focusing on Twitter for use in scholarship [24,29]. In academic literature, Twitter is often a topic of research and is believed not only to aid in diffusing innovation [35] but to be able to forecast the stock market [36], [37], deal with natural disasters like the earthquake in Haiti [38], or evaluate companies' brand presence [39]; Twitter can also simply serve as a news channel or an outlet for opinion sharing on consumer brands [40]. Twitter has been chosen out of other social networks for this research due to its structure, simplicity and power for dissemination of innovation (as described in Chang 2010 [35]).

Figure 1 represents the anatomy of a tweet and shows what data points would become available from a tweet.

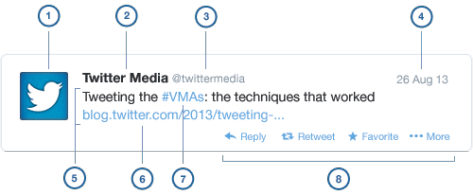


Fig. 1. Tweet anatomy

Profile picture: The image of the associated account.

Twitter account name: The name of the Twitter account. It might represent a real name of an individual, organization, entity or some other identification; the name may also be fake.

Twitter @username: @username is the unique identity on Twitter. The @ sign is also used to mention people in Tweets, like this: Hello @Twitter!

Tweet timestamp/date: This tells you when the Tweet was sent; the geo location of the tweet is also available here and is visible by clicking on the tweet itself.

Tweet text: Every tweet is fewer than 140 characters and may contain text as well as emoji.

Links: A tweet can carry a link to other websites, articles, photos and videos.

Hashtags: A hashtag is any word or phrase with the # symbol immediately in front of it. This symbol turns the word into a link that makes it easier to find and follow a conversation about that topic.

Tweet actions: An interaction in form of reply, retweet and favoring a Tweet.

The experiment has been conducted to collect the activity related to the startup ecosystem in each country, which is a good representation of societal practice of entrepreneurship. Startups are increasingly seen as significant contributors to national job-creation [41]; employment and gross national product data demonstrated the shift to an innovative startup-dominated economy [41]. Therefore, fostering the startup ecosystem is seen as the measure for improving national economy [42].

Based on surveying the literature and previous experiments, we took the assumption that startups are considered to be the main societal entrepreneurial activity. Twitter is a SNS platform which well represents and acts as support infrastructure for startups which are socially active. The study took the initiative to collect a sample of tweets from a region (country) and extract features (words and hashtags) related to startup activity; we have applied techniques to decompose hashtags, analyze them, and reuse the information extracted for classification purposes.

Data are collected using the Twitter Search API and the process of capturing relevant tweets for each country was benefited by the experiment done in Mohout et al 2011 [43] which was an attempt for constructing the innovation radar by utilizing Twitter data. The initial search for targeting the tweets which is oriented about startup activity was facilitated by looking at the associated keywords and hashtags aligned to each hashtag, as experimented in Mahout 2015 [43]. This process enabled us to collect and construct a searching query tailored for each country (by utilizing geo location of tweets to isolate the countries) regarding their startup activity discussion on Twitter. In the next step, we constructed the searching query in order to extract the number of tweets for a year for each study country. Figure 2 is an illustration of the step taken.

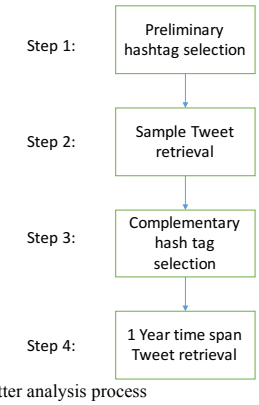


Fig. 2. Twitter analysis process

The result of this experiment is a proxy which illustrates the entrepreneurial activity for the studied countries. The numbers will be used in the output section among other metrics of the DEA model in order to calculate the efficiency ratio of countries in utilizing their social capital resources towards entrepreneurial-oriented activities.

The ratio indicating countries' entrepreneurial activity on Twitter is normalized by the population size and internet penetration rate in each country which was obtained from Internet Live Stats [3]. Figure 3 is a representation of the designed ratio in comparison to internet penetration. It illustrates the capacity achieved by a country regarding to the entrepreneurial oriented discussion in social media.

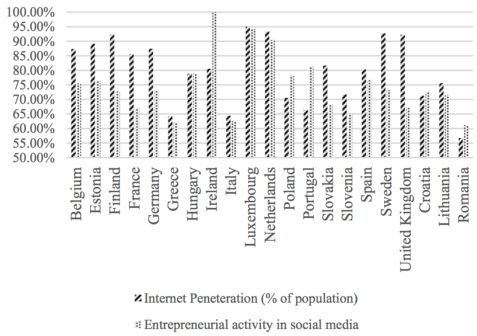


Figure 3 Capacity of entrepreneurial oriented activity in social media

IV. METHODOLOGY

The aim of the study is to assess the effectiveness and impact of social capital resources in boosting entrepreneurial-oriented activities. This study adopts both quantitative and qualitative approaches, depicting the complex issue of countries' entrepreneurial-oriented activity through indicators both traditional and new, which enables to a benchmarking practice on the analysis of the results. The following sections contain brief introductions to the rational for experimenting benchmarking practice as well as efficiency model, the required indicators and model construction to perform the analysis.

A. *Benchmarking*

The objective of benchmarking is to understand and evaluate the current position amongst competing peers on a particular objective in relation to best practice and to identify areas and means of performance improvement [44]. Benchmarking been used as a tool to improve business or organization' performance and competitiveness in business life, and has recently been adopted in both public and semi-public sectors [45]. While the general aspects of benchmarking include the evaluation and improvement of performance by learning from others, researchers have also begun investigating the scientific approaches to benchmarking, proceeding from practice towards theorizing [45].

The essence of performance analysis inherent in benchmarking has been leveraged in this study. This idea of a comparison between units of analysis by relative efficiency scores is found in the performance analysis literature [46]–[48]. Greiling 2006 [46], describes benchmarking as a process to make learning easier as a continuum and systematic procedure of measuring products, services, and practices aiming at correcting failures and improving outcomes.

One of the most commonly used benchmarking frontier techniques is Data Envelopment Analysis (DEA). Efficiency scores among studied countries will be projected and ranked accordingly to get a concise clear of the highest efficient cases.

B. *Data envelopment analysis*

The Data Envelopment Analysis (DEA) methodology was first introduced by Charnes, Cooper and Rhodes in 1978 [32] and then extended by Banker, Charnes and Cooper in 1984 [49] This

[3] *Internet live stats: Elaboration of data by International Telecommunication Union (ITU) and United Nations Population Division accessible from: http://www.internetlivestats.com/internet-users-by-country/*

methodology has been widely used for estimating technical efficiencies of Decisions Making Units (DMU). According to G. Tavares, in the period of 1978 to 2001 alone, there have been more than 3600 papers, books, and other written documents by more than 1600 authors related to DEA and the numbers are ever growing [50]. DEA is a mathematical programming method that provides a single measure of efficiency. It is calculated with information about the use of multiple inputs and multiple outputs and results in a score, called a frontier in these analyses, which represents the best practice. From this best efficiency frontier, the relative efficiency of DMUs is calculated. For each DMU, DEA presents an efficiency score, typically ranging between zero (inefficient) and 1 (efficient), which indicates inefficiencies. Furthermore, the DEA efficiency frontier can be used as a guideline so that inefficient DMUs can improve their inputs and outputs and reach the efficiency frontier. The maximization of the efficiency ratio will be achieved by solving a linear programming problem. The mathematical formula represented in the following known as "CCR" has been adopted from Cooper et al. [50] which is named after the creators Charnes, Cooper and Rhodes.

$$
\begin{aligned}
&\text{maximise} && \theta = \sum_{r=1}^{s} \mu_r y_{ro} \\
&\text{subject to} && \sum_{i=1}^{m} v_i x_{io} = 1 \\
(1) \quad & && \sum_{r=1}^{s} \mu_s y_{rj} - \sum_{i=1}^{m} v_m x_{mj} \le 0 \\
& && (j = 1, \cdots, n) \\
& && v_1, v_2, \cdots, v_m \ge 0 \\
& && \mu_1, \mu_2, \cdots, \mu_s \ge 0
\end{aligned}
$$

Where

$\theta =$ sum of virtual outputs

$x_{ij} =$ amount of input i used by DMU j

$y_{rj} =$ amount of output r produced by DMU j

$v_i =$ weight of input i

$\mu_r =$ weight of output r

$n =$ number of DMUs

Subscript o refers to the DMU whose efficiency is calculated.

The model is solved n times to determine the relative efficiency for each DMU.

The CCR-model is sometimes referred to as the CRS-model because it builds on the assumption of constant returns to scale (CRS). Constant returns to scale mean that outputs increase in direct relation to an increase in the inputs, or similarly decreases in inputs bring about relative decreases in outputs. However, different return to scale assumptions may have different impacts on the allocation and besides, the homogeneity among DMUs also needs to take into account. Therefore, it is necessary to consider the problem under the variable return to scale (VRS) assumption (called BCC model) [49]. Mathematically, the BCC linear programming model may be represented as follows [51]:

$$
\begin{aligned}
&\text{maximize:} && \sum_{r=1}^{s} u_r y_{ro} - u_o \\
&\text{subject to:} && \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - u_o \le 0 \\
& && \sum_{i=1}^{m} v_i x_{io} = 1 \\
& && -u_r \le -\varepsilon \\
& && -v_i \le -\varepsilon
\end{aligned}
$$

The VRS quality of the model makes it more flexible and less strict than the previous CCR-model. As a rule, CCR-efficiency scores never exceed BCC-scores, although the opposite often is true. The calculation in the study will present both CCR scores and BCC scores.

1) CONSTRUCTION OF THE DEA-MODEL

A model tends to provide a simplification of reality to promote better understanding. In developing models a trade of is with the selection of variables which are capturing the reality and might be accused for an objectivity of viewer's perception. This section tries to bring transparency to the procedure of constructing the DEA-model. It will present the simplification of the efforts on building capabilities in social capital as an initial resource towards entrepreneurial-oriented environment at a country level. Based on the discussion in the literature review section on the social capital phenomenon and its attributes, which emphasizes on skills and knowledge and education, we engage in a process of converting the attributes into existing variables. The output variables were purposefully selected to challenge the efficiency of the capacities seen in social capital. Therefore in regards to the context of this study the variables selected to represent entrepreneurial capacity within society. In

Figure 4 an attempt is a simplified sketch of the model.

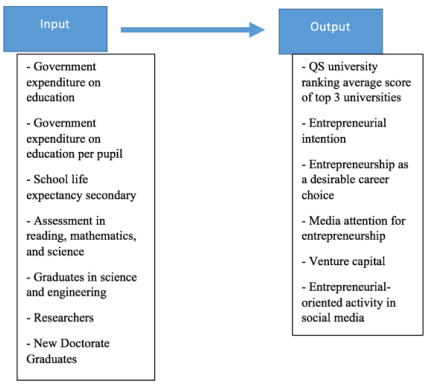


Fig. 4. DEA Input and Output model

The analysis is based on an extensive data set which is modified for applicability of the DEA-model. European countries was the study subject in which 22 countries proceeded to analysis due to restriction of DEA model with handling missing value. Country is the level of analysis due to greater availability of data comparing to regions and is a sensible level due to language and cultural barrier for the context of our study. Each variable is normalized by the size of the representing country as the intention of the analysis is to challenge the efficiency within the capacity provided by the country.

The quantification of inputs and outputs is presented in Tables 1 with the description and the sources of the variables. The selected variables comprise 7 input and 6 output, all of which are relative measures to accommodate for differences in country size. We allow the model to use a one-year time lag between inputs and outputs. Therefore, the input data are from year 2015 and output data are from year 2016.

Table 1. Variable discription

Input factors	**Proxy Indicator**	**Source of variable**
Government expenditure on education (% of GDP)	Government operating expenditures in education, including wages and salaries and excluding capital investments in buildings and equipment, as a percentage of gross domestic product (GDP).	UNESCO Institute for Statistics, UIS online database
Government expenditure per pupil, secondary (% of GDP per capita)	Government spending on education divided by the total number of secondary students, as a percentage of GDP per capita. Government expenditure (current and capital) includes government spending on educational institutions (both public and private), education administration, and subsidies for private entities (students/households and other private entities).	UNESCO Institute for Statistics, UIS online database
School life expectancy, primary to tertiary education (years)	Total number of years of schooling that a child of a certain age can expect to receive in the future, assuming that the probability of his or her being enrolled in school at any particular age is equal to the current enrolment ratio for that age.	UNESCO Institute for Statistics, UIS online database
Assessment in reading, mathematics, and science	The Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA) develops threeyearly surveys that examine 15-yearold students' performance in reading, mathematics, and science. The scores are calculated in each year so that the mean is 500 and the standard deviation 100. The scores for China come from Shanghai; those for India from Himachal Pradesh	OECD Programme for International Student Assessment (PISA)

	and Tamil Nadu (average); those for the United Arab Emirates from Dubai; and those for the Bolivarian Republic of Venezuela from Miranda. These scores are those from the GII 2015 report.	
Tertiary graduates in science, engineering, manufacturing, and construction (% of total tertiary graduates)	The share of all tertiary graduates in science, manufacturing, engineering, and construction over all tertiary graduates.	UNESCO Institute for Statistics, UIS online database
Researchers, full-time equivalence (FTE) (Normalized score)	Researchers per million population, fulltime equivalence. Researchers in R&D are professionals engaged in the conception or creation of new knowledge, products, processes, methods, or systems and in the management of the projects concerned. Postgraduate PhD students (ISCED97 level 6) engaged in R&D are included.	UNESCO Institute for Statistics, UIS online database
New Doctorate Graduates (Normalized score)	New Doctorate Graduates (ISCED 6) per 1000 population aged 25-34	Innovation Union scoreboard
Output factors		
QS university ranking average score of top 3 universities	Average score of the top three universities per country. If fewer than three universities are listed in the QS ranking of the global top 700 universities, the sum of the scores of the listed universities is divided by three, thus implying a score of zero for the non-listed universities.	QS Quacquarelli Symonds Ltd, QS World University Ranking 2015/2016, Top Universities.

Entrepreneurial intention	Percentage of 18-64 population (individuals involved in any stage of entrepreneurial activity excluded) who intend to start a business within three years	Global Entrepreneurship Monitor
Entrepreneurship as a desirable career choice	Percentage of 18-64 population who agree with the statement that in their country, most people consider starting a business as a desirable career choice.	Global Entrepreneurship Monitor
Media attention for entrepreneurship	Percentage of 18-64 population who agree with the statement that in their country, you will often see stories in the public media about successful new businesses	Global Entrepreneurship Monitor
Venture capital investment	early stage, expansion and replacement as % of GDP	Global Entrepreneurship Monitor
Entrepreneurial-oriented activity in social network services	A ratio representing each country discussion in social media regarding entrepreneurial-oriented activities	Twitter

Following the variable selection and definitions, Table 2 is a descriptive statistic of the variables which will be used in the DEA model.

Table 2 Descriptive statistics

	Mean	*Median*	*Std. Deviation*	*Min*	*Max*
GOV-EXP-EDU	73.46667	89.4	33.48134	4.2	95.7
GOV-EXP-EDU-PUP	8.809524	5.9	7.376985	4.1	33.6
SCHLIF	4573.671	4893.2	2925.147	14.1	10678.8
AS-RED	92.45714	2.2	186.9299	0.7	502.5
GRAD-SCI	7.647619	4.9	6.543599	3	22.3
RESEARCHER	3522.038	3438	1492.296	862	7223.3
DOCTOR	0.572286	0.526	0.263068	0.144	1
QSRANK	9.738095	4.4	20.74855	0	98.9
ENTRE-INTEN	12.51238	11.36	5.787126	5.93	31.7
ENTRE-CARER	56.94286	55.56	9.750038	40.66	79.11

MEDIA-ENTRE	52.7	51.41	11.18362	33.47	75.68
SOCIAL-MEDIA	74.72857	73.2	10.01233	61.2	100
VC-DEAL	0.057974	0.054078	0.036789	0.00093	0.136353

V. RESULTS AND DISCUSSION

The result of benchmarking and efficiency assessment has been facilitated with MaxDEA software. MaxDEA basic edition developed by Cheng, G., and Z. Qian [52] has been used to determine efficiency ratios. The comprehension of efficiency ratio should be determined relative to other DMUs. It has been recommended to not to rely solely on the results of DEA as an individual analysis, but it is equally important to understand its significance in supporting decision-making when properly interpreted in conjunction with other information [13]. The relative notion of the efficiency score would interpret in a way that countries would be compared to each other that share a similar mix of inputs, which translates to similar possibilities of being efficient. The main objectives of DEA is to measure the efficiency of a DMUs by a scalar measure ranging between zero (the worst) and one (the best). The efficient state of DMU in DEA analysis is described in which further improvement won't achieve without harming some other input or output. Therefore, in the context of this study, countries are turn out efficient whom utilized their capacity productively than less efficient ones.

The efficiency scores are estimated using a variable return to scale, output oriented DEA with 7 inputs and 6 output. Decision making units (DMUs) with efficiency scores equal to 1 are efficient, while DMUs with scores greater or less than 1 are inefficient. Meanwhile it is important to note that the efficiency scores are relative scores. High relative efficiency scores do not mean that there is no room for improving performance. The results of these evaluations are displayed below in table 3.

Table 3. DEA analysis BCC and CCR efficiency scores

BCC Output DMU	*Score*	*CCR output DMU*	*Score*
Ireland	1	Ireland	1
Netherlands	1	Netherlands	1
Luxembourg	1	Luxembourg	1
Hungary	1	Hungary	1
Portugal	1	Portugal	1
Poland	1	Poland	1
Slovakia	1	Slovakia	1
Bulgaria	1	Bulgaria	1
Latvia	1	Latvia	1
Romania	1	Romania	1
United Kingdom	1	Croatia	0.998495
Finland	1	Greece	0.99245
Croatia	1	Italy	0.96797
Greece	1	Finland	0.959859
Slovenia	0.995086	United Kingdom	0.954028
Italy	0.969573	Belgium	0.943379
Germany	0.955024	Slovenia	0.92143
Belgium	0.946866	Germany	0.905398
Sweden	0.90235	Sweden	0.879499
Estonia	0.868719	Estonia	0.829102
Spain	0.723702	Spain	0.706288

We have also introduced a variable to represent the DMU's performance regarding the venture capital raised funds among the output variables. The inclusion of VC variable was meant to see how the social capital in terms of education has been translated to efficiency on venture capital investments as well as entrepreneurial-oriented activity. The results of these evaluations are displayed in Table 4.

Table 4. DEA analysis with VC variable. BCC and CCR efficieny scores

BCC Output DMU	*VC Score*	*CCR DMU*	*VC Score*
Finland	1	Finland	1
Ireland	1	Ireland	1
United Kingdom	1	United Kingdom	1
Netherlands	1	Netherlands	1
Luxembourg	1	Luxembourg	1
Portugal	1	Portugal	1
Hungary	1	Hungary	1
Estonia	1	Estonia	1
Poland	1	Poland	1
Romania	1	Romania	1
Latvia	1	Latvia	1
Bulgaria	1	Bulgaria	1
Slovakia	1	Slovakia	1
Sweden	1	Croatia	0.999518
Croatia	1	Greece	0.99245
Greece	1	Italy	0.96797
Slovenia	0.995086	Sweden	0.962335
Italy	0.969573	Belgium	0.943379
Germany	0.955024	Germany	0.935738
Belgium	0.946866	Slovenia	0.92143
Spain	0.743045	Spain	0.728821

VI. Discussion and conclusion

The study aimed to achieve meaningful setup in order to communicate the results in which countries position establishes a base on their efficiency status on utilizing their capacities. The communication of the countries position in scale was facilitated by ranking in which the relationship between set of items are simplified into sequence of ordinal numbers. Rankings make it possible to evaluate complex information according to certain criteria. Therefore, utilizing the DEA model considered necessary to construct and calculate the countries ranking.

Societal and cultural capacity have been recognized as important factors for innovation-driven economies and have been calculated in different forms such as human capital in Global Innovation Index 2014 [53]. This paper attempted to capture the social capital educational capacity and illustrate its indirect effect to the entrepreneurship-oriented activities.

Entrepreneurship-oriented activities are triggered by many factors, and it is good to distinguish between "necessity entrepreneurs" and "opportunity entrepreneurs". That sort of entrepreneurs who are the most valuable are the opportunity entrepreneurs, who are the new entrepreneurs who are starting their own businesses because they see opportunity, not because they are out of work and unable to get a job. Therefore, distinguishing of this two groups is important and leads to a better understanding of the overall capacity of a society for introducing opportunity entrepreneurs.

The application of Data Envelopment Analysis is proposed as a methodology to overcome the problems related to the lack of methodology to assign the correct weightings for the calculation of indexes and to the subjectivity of the interpretations of results [54], [13]. The interpretability of efficient scores are due to the benchmarking feature which is based on indicators that enable decision makers to know the relative position of a DMU with respect to others. The benchmarking results utilized for the ranking lists in Tables 3 and 4, based on the BCC and CCR efficiency scores of each country.

The rankings resulted in CCR calculation had similar or lower efficiency scores compering to BCC scores. The indication of this difference in ranking between the CCR and BBC suggest that some countries disadvantaged by the shift to constant return to scale assumption. However, almost half of the countries obtained the efficient status such as Ireland, Netherlands, Luxembourg, Hungary, Portugal, Poland, Slovakia, Bulgaria, Latvia and Romania.

It is important to note that the reason to overvalue poor countries can be described since the model doesn't encounter economic indicators. The intention for this research was to evaluate the social capacity and its transformation to entrepreneurial oriented activity. Therefore, the model may confuse prowess in economic development with the ability of entrepreneurial-oriented activities. In summary, DEA has potential to become a meaningful analysis tool for evaluating the achievement of higher capacity in social capital at the country level towards entrepreneurial-oriented activity.

The study where able to add a new perspective on the established ranking system by proposing efficiency rather than proficiency. Non-parametric techniques such as DEA model can be applied to panel data such as this practice to shed light on changes in efficiency over time. The context of the study was relevant and important as it focuses on the social capital capacity and the efficiency of its utilization in an economy leading towards more entrepreneurial activity. The study had also discussed the importance of social network services outlets as the pulses of society and their dominant nature in hosting big portion of discussion. Furthermore, the SNS data had been leveraged in order to capture the vibe and intensity of the discussion on the SNS platform about startup and entrepreneurial activity. The result of the attempt was translated into a metric which was used in the DEA model. Therefore, the study and constructed model has the unique characteristic of examining the efficiency of countries in utilizing their social capital resources towards entrepreneurial-oriented activity. The efficient countries, according to the model, had a better balance in regards to their social capital and the entrepreneurial-oriented activity in society when compared to less efficient ones.

In order to address the limitation of the study, the choices for the theoretical framework (e.g. definition of human capital, social capital and entrepreneurial activity) have an impact on the choice of variables, which also narrows the scope of research and the results to be achieved. Although

the Data Envelopment Analysis showed a great deal of freedom on scale and quality of data which can be used but also imposes certain restrictions such as sensitivity of the results to the selection of inputs and outputs and the lack of possibility for testing the specifications.

REFERENCES

[1] Å. L. Dahlstrand and L. Stevenson, "Innovative entrepreneurship policy: linking innovation and entrepreneurship in a European context," *Ann. Innov. Entrep.*, vol. 1, pp. 1–15, 2010.

[2] P. F. Drucker, "Managing for the future," *Routledge*, no. 1, p. 281, 1993.

[3] P. a. Geroski, J. Mata, and P. Portugal, "Founding conditions and the survival of new firms," *Strateg. Manag. J.*, vol. 31, pp. 510–529, 2010.

[4] R. S. Ralph H. Kilmann, Mary J. Saxton, *Gaining Control of the Corporate Culture*, vol. 32, no. 3. 1987.

[5] E. G. Carayannis and D. F. J. Campbell, "'Mode 3' and 'Quadruple Helix': toward a 21st century fractal innovation ecosystem," *Int. J. Technol. Manag.*, vol. 46, p. 201, 2009.

[6] R. Florida, *The Rise of the Creative Class--Revisited: Revised and Expanded*. 2014.

[7] Statista, "Number of social media users worldwide from 2010 to 2020 (in billions)," 2016. [Online]. Available: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/. [Accessed: 01-Nov-2016].

[8] C. Anderson and M. Wolff, "The Web Is Dead. Long Live the Internet," *WIRED*, Aug-10AD.

[9] Internet Scociety, "Global Internet Report," 2014.

[10] M. G. Russell and K. Still, "Engines driving knowledge-based technology transfer in business incubators and their companies," *Proc. 32nd Annu. Hawaii Int. Conf. Syst. Sci.*, p. 9, 1999.

[11] H. Etzkowitz and L. Leydesdorff, "The dynamics of innovation: from National Systems and Mode 2'' to a Triple Helix of university–industry–government relations," *Res. Policy*, vol. 29, pp. 109–123, 2000.

[12] European Commission, "Research and Innovation performance in EU Member States and Associated countries," p. 334, 2013.

[13] A. Kutvonen, "Ranking Regional Innovation Policies: Dea-Based Benchmarking in a European Setting," 2007.

[14] Oxford, "Human capital," *Oxford*. [Online]. Available: https://en.oxforddictionaries.com/definition/human_capital. [Accessed: 01-Feb-2017].

[15] "Human Capital," *BusinessDictionary*. [Online]. Available: http://www.businessdictionary.com/definition/human-capital.html. [Accessed: 02-Mar-2017].

[16] N. Bontis, W. C. C. W. C. C. Keow, and S. Richardson, "Intellectual capital and business performance in Malaysian industries," *J. Intellect. Cap.*, vol. 1, no. 1, pp. 85–100, 2000.

[17] R. H. Stevens, "Managing human capital: How to use knowledge management to transfer knowledge in today's multi-generational workforce," *Int. Bus. Res.*, vol. 3, no. 3, pp. 77–83, 2010.

[18] N. Bontis, "Intellectual capital: an exploratory study that develops measures and models," *Manag. Decis.*, vol. 36, no. 2, pp. 63–76, Mar. 1998.

[19] L. Edvinsson and M. Malone, "Intellectual capital: realizingyour company's true valor by finding its hidden brainpower," 1997.

[20] M. A. Hitt, "The Essence of Strategic Leadership : Managing Human and Social Capital," *Leadership*, vol. 9, no. 1, 2003.

[21] T. S. Bateman and J. M. Crant, "The proactive component of organizational behavior: A measure and correlates," *J. Organ. Behav.*, vol. 14, no. 2, pp. 103–118, 1993.

[22] A. Smith, *Wealth of Nations*. 1892.

[23] A. Smith, "The Wealth of Nations - An inquiry into the Nature and Causes of the Wealth of Nations," p. 572, 1993.

[24] G. Becker, "Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education," *New York*, 1971.

[25] S. E. H. Jensen, T. N. Rasmussen, and T. F. Rutherford, "Economic Transition, Entrepreneurial Capacity, and Intergenerational Distribution," p. pages, 2002.

[26] R. Nowak, "Entrepreneurial Capacity and Culture of Innovation in The Context of Opportunity Exploitation," Urbana, Illinois, 2014.

[27] G. S. Becker, "Investment in Human Capital: A Theoretical Analysis," *J. Polit. Econ.*, vol. 70, no. 5, Part 2, pp. 9–49, Oct. 1962.

[28] J. S. Coleman, "Social Capital in the Creation of Human Capital," *Am. J. Sociol.*, vol. 94, no. 1988, p. S95, 1988.

[29] C. Steinfield, N. B. Ellison, and C. Lampe, "Social capital, self-esteem, and use of online social network sites: A longitudinal analysis," *J. Appl. Dev. Psychol.*, vol. 29, no. 6, 2008.

[30] C. Steinfield, J. M. DiMicco, N. B. Ellison, and

C. Lampe, "Bowling Online : Social Networking and Social Capital within the Organization," *Distribution*, pp. 245–254, 2009.

[31] J.-H. Lin, W. Peng, M. Kim, S. Y. Kim, and R. LaRose, "Social networking and adjustments among international students," *New Media Soc.*, vol. 14, no. 3, pp. 421–440, May 2012.

[32] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *Eur. J. Oper. Res.*, vol. 2, no. 6, pp. 429–444, 1978.

[33] J. Priem and B. M. Hemminger, "Scientometrics 2.0: Toward new metrics of scholarly impact on the social web," *First Monday*, vol. 15, no. 7, 2010.

[34] X. L. . M. T. . D. Giustini, "Validating online reference managers for scholarly impact measurement," in *ISSI 2011 Conference*, 2011.

[35] H. C. Chang, "A new perspective on Twitter hashtag use: Diffusion of innovation theory," *Proc. ASIST Annu. Meet.*, vol. 47, 2010.

[36] X. Zhang, H. Fuehres, and P. A. Gloor, "Predicting Stock Market Indicators Through Twitter 'I hope it is not as bad as I fear,'" *Procedia - Soc. Behav. Sci.*, vol. 26, no. 2007, pp. 55–62, 2011.

[37] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, 2011.

[38] S. Muralidharan, L. Rasmussen, D. Patterson, and J. H. Shin, "Hope for Haiti: An analysis of Facebook and Twitter usage during the earthquake relief efforts," *Public Relat. Rev.*, vol. 37, no. 2, pp. 175–177, 2011.

[39] A. Hajikhani, J. Porras, and H. Melkas, "Brand Analysis in Social Network Services: Results from Content Analysis in Twitter Regarding the US Smartphone Market," *Int. J. Innov. Technol. Manag.*, p. 1740008, Oct. 2016.

[40] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, no. 11, pp. 2169–2188, 2009.

[41] J. E. Sohl, "Angel investing: Changing strategies during volatile times," *J. Entrep. Financ. Bus. Ventur.*, vol. 11, no. 2, pp. 27–47, 2006.

[42] R. L. La Rovere, L. de M. Ozorio, and L. de Jesus Melo eds., "Entrepreneurship in BRICS: Policy and Research to Support Entrepreneurs," p. ix, 2015.

[43] O. Mohout and I. Fiegenbaum, "The power of Twitter : Building an innovation radar using social media," no. June, 2015.

[44] J. Tidd and J. Bessant, *Managing Innovation for Growth in High Technology Small Firms*. 2005.

[45] P. Kyrö, "Revising the concept and forms of benchmarking," *Benchmarking An Int. J.*, vol. 10, no. 3, pp. 210–225, 2003.

[46] D. Greiling, "Performance measurement : a remedy for increasing the efficiency of public services ?," *Int. J. Product. Perform. Meas.*, vol. 55, no. 6, pp. 488–465, 2006.

[47] S. C. Selden and J. E. Sowa, "Testing a Multi-Dimensional Model of Organizational Performance: Prospects and Problems," 2003.

[48] W. R. Scott and F. Gerald, "Organizations and Organizing: Rational, Natural, and Open System Perspectives," *Organ. Organ.*, p. 452, 2007.

[49] R. D. Banker, A. Charnes, and W. W. Cooper, "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Manage. Sci.*, vol. 30, no. 9, pp. 1078–1092, 1984.

[50] W. Cooper, L. Seiford, and K. Tone, *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. 2007.

[51] W. F. Bowlin, "Measuring Performance: An Introduction to Data Envelopment Analysis," *J. Cost Anal.*, vol. 7, pp. 3–27, 1998.

[52] G. Cheng and Z. Qian, "MaxDEA Linear Programming, Version 5.2 (computer) programme." 2013.

[53] R. Scott and S. Vincent-Lancrin, "The Global Innovation Index 2014: The Human Factor In Innovation," 2014.

[54] J. A. Rodriguez Diaz, E. Camacho Poyato, and R. Lopez Luque, "Applying benchmarking and data envelopment analysis (DEA) techniques to irrigation districts in Spain," *Irrig. Drain.*, vol. 53, no. 2, pp. 135–143, 2004.

Publication V

Hajikhani, A., Silva, M., Porras, J.

Crowd Intelligence Participation in Digital Ecosystem: Systematic Process for Driving Insight from Social Network Services Data

Reprinted with permission
In Proceedings of the Academy of Management Specialized Conference on Big Data and Managing in a Digital Economy 2018

Crowd Intelligence Participation in Digital Ecosystem: Systematic Process for Driving Insight from Social Network Services Data

ABSTRACT

In the increasingly fast pace of digitalization, Social Network Services (SNS) are acting as a major component in the transformation process. Meanwhile, the progress of computational power and data analytics techniques necessitate revisiting the massive data generated from SNS for accurate insights for better decision making. In this article, we propose a systematic process for analyzing SNS data utilizing the advancement in text analytics and topic modeling. Furthermore, the proposed methodology derives insight from crowd intelligence contributions on Twitter regarding the sample case of Fukushima incident. The relevant tweets were retrieved, preprocessed, and textually analyzed to reveal the topic evolutionary pattern in the discussions. The topical analysis and visualization indicates more coherent and less topic proximity in discussion over time.

Keywords: Digital Platform; Digitalization; Social Network Services; Content Analysis; Online Communication

Introduction

Digitalization is the process of integrating digital technology into everyday life, and is happening with an increasing trend. In an increasingly digital economy, data is playing a major role in digitalization and transformation of routines towards more integration with digital technologies. The convergence of extremely large data sets known as "big data," which are generated by technologies such as social media, mobile, analytics, and cloud computing, has led to an unprecedented wave of digitalization that is currently fueling innovation in businesses and society (Legner et al., 2017). Modern science has become collaborative and digital. The Internet has supported the emergence of scientific digital platforms that globally connect programmers and users of novel digital scientific products such as scientific interactive software tools (Brunswicker, Matei, Zentner, Zentner, & Klimeck, 2017).

In the face of a digital revolution, national and regional governments are increasingly defining digitalization as a strategic priority and are setting up large-scale initiatives to foster digital transformation of science, industry, and society (European commision, 2016; OECD, 2016). As digitalization matures at various fronts of economies and societies, participation in digital ecosystems and platforms is increasing, along with the reach and range of these networks. Navigating through this digital ecosystem has provided a huge number of opportunities while introducing challenges as well. ICT have become widely available to the general public, both in terms of accessibility as well as cost; as a result, widespread and

affordable broadband access is one of the means of promoting a knowledge-based and informed society (Eurostat, 2017). The fast progress of technology has resulted in societies and communities that are connected and maintaining their communication on platforms such as social media or Social Network Services (SNSs). A huge amount of data is generated in SNSs so it raises the major research questions as follow:

- How can the role of Social Network Services (SNSs) as a major component in digital platforms be materialized?
- How does the collective knowledge in the process of generation and dissemination in SNSs evolve in structure and coherency?

We review the emerging digital platform phenomena and scrutinize the theoretical perspectives, research methods, and data for better understanding of the concept. Next, we obtain a system perspective framework to draw out the important elements of the digital platform. After reviewing the digital platforms' historical roots and specifications, we then focus on newly emerging entities such as SNSs and their impact and presence in digital platforms. Furthermore, we present ways and technologies in which SNS data can be analyzed for better insights. A systematic process will be proposed for SNS data analytics and a case study will be employed to put the methodology and proposed computational techniques into practice. The other end goal of our research is to better understand process and dynamics of knowledge creation among communities in online discussions.

Digital Platforms

Digital platforms bring together two distinct streams of research. The emergence of digital technologies provides new avenues that allow the combination of material properties and digital properties (i.e., hardware and software) to create artefacts that are both flexible and reprogrammable (Kallinikos, Aaltonen, & Marton, 2013; Yoo & Henfridsson, 2010). According to Gawer (2014), platform considers as a value creation mechanism for its stakeholders that is built around a core and is stable over time. Various conceptualizations of digital platforms exist due to the distributed nature of digitalization (Henfridsson, Mathiassen, & Svahn, 2014). Therefore, it is necessary to elaborate on the two distinct stream of literature from which it borrows.

A platform is known as the structure which bring together multiple user groups for the adoption of technology. Increasing adoption levels can trigger positive feedback cycles that further increase the usefulness of the technology (Arthur, 1989). Due to the network effect, a technology's adoption increases by growth of the users (Katz & Shapiro, 1985; Shapiro & Varian, 1998). Network effects or network externalities will enrich platforms with complementary services, communities of users, higher-quality products, and new market opportunities (Dew & Read, 2007). Examples of network effects are social network services, which become more valuable if more end-users join the platform. Platforms are closely related to ecosystems and therefore the two terminologies has been used interchangeably (de Reuver, Sørensen, & Basole, 2017; Hajikhani, 2018). Both platforms and ecosystems are

often associated with “network effects” - that is, as the network grows and more users join, the more valuable the platform becomes to the owner and to the users themselves, due to increased access to a growing network of users and complementary innovations (Gawer & Cusumano, 2014). While ecosystems focus on structure and interdependence, platforms are concerned with governance. Platforms, explains by Adner (2017), “hold a hub position in a network of interactions” and “exercise power through centrality”.

While the management literature on platforms is focused on modularization and governance of components in a hierarchical design (Clark, 1985), digital platforms imply distributedness and homogenization of data (Kallinikos et al., 2013; Yoo & Henfridsson, 2010). Such characteristics of digitalization can lead to multiple inheritance in large distributed technical arrangements, meaning there is no single owner that owns the platform core and dictates its design hierarchy (Henfridsson et al., 2014). With the increasing availability of digital data about the digital ecosystem, its components and the relationships between them, we have an increased ability to apply data driven analysis and visualization approaches to generate novel insights into ecosystems and the role of platforms (de Reuver et al., 2017). Digital platforms support new ways of interacting within communities and through mediated co-creation. They allow ordinary citizens to share their thoughts while consuming others’ content. Therefore, a digital platform can be characterized as a sociotechnical assemblage encompassing the technical elements (of software and hardware) and associated organizational processes and standards (Tilson, Sørensen, & Lyytinen, 2012).

The challenges for materializing digital platforms has been mentioned as comparability of research units and unit of the analysis (Tilson, Lyytinen, & Sørensen, 2010). In addition, live stream data and large accessibility to data require efficient tools so to encounter the dynamics of digital platforms and ecosystems within a time horizon (Tilson et al., 2010). The issue of how to govern digital platforms is another continuing challenge to note. There are various concerns narrated by De Reuver (2017) with applying digital platform insights into practice, including questions such as how can scholars and practitioners effectively manage the intense velocity and scale at which data on digital platforms is generated? And how to develop computational capabilities and insights that allow greater understanding of changes in the platform and the resulting impact on platform components?

The first step to comprehend the digital platform or ecosystem is to think of the digital world as a *system* of interconnected elements that is (coherently) organized in a way that achieves a goal (Meadows, 2011). According to Meadows (2011) the three key concepts in this basic definition of system are *components*, the *interconnections*, and a *function* or *purpose*.

- *Components* are the things that make up the system. Components don't always have to be physical and tangible; they can also be intangible things that influence the behavior of the system.

- *Interconnections* are the relationships that hold the elements of the system together. Information flows are another example of interconnection, and are critical for many systems, especially digital ones.
- *Function or Purpose* is the state which the system is supposed to accomplish.

In this study, we are pursuing a path for focusing on one of the major components of digital platforms known as Social Network Services (SNSs). The process of understanding SNSs will proceeded by visiting the recent advanced methodological capabilities from data analytics, machine learning and visualization. We will continue our investigations with an empirical analysis of the dynamics of the discussions in the Twitter as one of the popular SNSs.

Social Network Services Role in Digital Platforms

Every day, millions of users worldwide are connected and receive their news via online social networks, warranting researchers to study the mechanisms behind human interactions (Anderson & Caumont, 2014). The influence of social media on political and social issues is getting greater and greater (Eom, Puliga, Smailovic, Mozetic, & Caldarelli, 2015). Internet penetration facilitates interaction in communication services such as email, chat, and messaging where this evolutionary path resulted in the domination of Social Network Services (SNSs) in today's information exchange. In other words, with the advent

of user-generated content and sharing features, SNS platforms are the new web experience. SNSs have been approached from various research disciplines as well as organizational studies where electronic networks were discussed as a way to create linkages to external knowledge resources. The cornerstone of discussion within electronic networks was the possibility to share information quickly, globally, and with large numbers of individuals to facilitate knowledge exchange (Powell, Koput, & Smith-Doerr, 1996). In the recent assessment made by the World Economic Forum (2017), "social media" is recognized as one of the forces driving transformational change across economies, industries, and global issues. SNS platforms enabled a shift in the way we communicate with each other. In this way, interpersonal interactions are no longer limited by time and space, and can better enable people to find and connect with others who share interests and beliefs as they extend relationships beyond the physical world. As one of the components of the digital platform, the emergence of SNSs has tremendous effects in our societies which yet has not received a great deal of attention for research. Figure 1 is a visualization the shows the positioning of social media interconnections and interdependencies between other important recognized topics according to the World Economic Forum assessment.

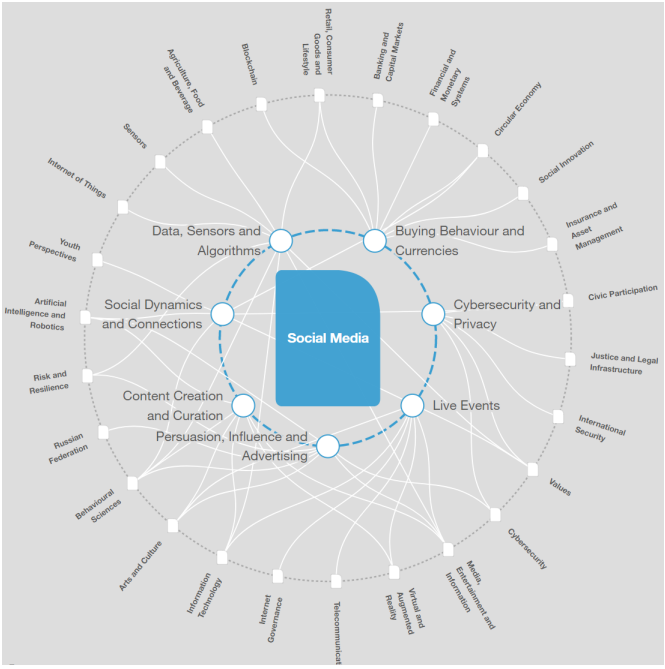


Figure 1. Social media interlinked positioning among other global challenges

(Adopted from World Economic Forum Global Trends 2017)

Among the key issues interacting in social media are content creation and curation, social dynamics, and live events. Social media has been a dominant venue where people either participate in or passively consume live events as they unfold. Recently, Twitter has been used for spreading news and updates around the world and has been shown to have application in emergency situations of natural disasters such as earthquakes, floods, hurricanes, and wildfires (Hughes & Palen, 2009; Kireyev, Palen, & Anderson, 2009; Muralidharan, Rasmussen, Patterson, & Shin, 2011; Starbird, Palen, Hughes, & Vieweg,

2010; Vieweg, Hughes, Starbird, & Palen, 2010). Social media's technology platforms allow for multidirectional network communication which can aid officials during disasters to compile a list of the injured and deceased, and contact family and friends of victims all while connecting and organizing both casualties and responders (Cooper, Yeager, Burkle, & Subbarao, 2015). Twitter has shown to have the potential to increase survival during tornado-related disasters (Lindsay, 2011). It is investigated that social media and SNSs have been increasingly used for building and supporting communities and affording self-expression and identity construction for individuals in the communities they belong to (Jaeger et al., 2007). The capabilities in SNSs have been leveraged to initiate real-time information network powered by communities and authorities.

Understanding the role of SNSs as one of the major components can give a better idea of the digital platform. SNSs represent contexts with new responsibilities for both academics and practitioners. It is necessary to recognize that these digitally enabled mass communication platforms are imposing fundamental shifts in how we understand people, society, and technology. The social infrastructure, such as intellectual and social capital, presented by SNSs is an indispensable endowment to the digital platforms as it allows "connecting people and creating relationships" (Albino, Berardi, & Dangelico, 2015). ICTs also offer new avenues for openness by providing access to SNS content and interactions that are created through the social interaction of users via highly accessibly Web-based

technologies. SNSs can be used to refer to both the enabling tools and technology and to the content that is generated by them.

As SNSs are an integral part of the mentioned applications, the motivation of this research is to drive insightful information from SNSs. In the next section, we will propose a systematic way for deriving insight from SNSs data so to better understand the unification and cohesion in SNS discussions. Furthermore, for better illustration of the systematic approach in SNS data, we run an experiment regarding the Fukushima incident. The systematic SNS data analysis will be replicated in order to collect data from Twitter so to observe the topical evolution of the discussion over time.

Utilizing Social Network Services Data to Tackle Social Environment Challenges

On March 2011, Japan's coast was hit by a 9.0 Mw earthquake which caused a nuclear energy accident in the Fukushima power plant. This accident went down in history as the second major event of its kind in the world and its effects will stand decades to come. During massive disasters such as this one, people rely on SNSs to get information (Hirschburg, 1986; Jung, 2012; Kim, Jung, Cohen, & Ball-Rokeach, 2004; Lowrey, 2004). At that time, SNSs were already playing an important role in the daily lives of many people around the world. As a result, websites dedicated to explaining radiation and natural events saw a significant spike in followers and interactions on Twitter and Facebook. The citizens' activity on SNSs was such that the prime minister's office decided to then to create a Twitter account dedicated to disaster management. The following weeks showed that the activity on social media was

not ceasing due to a growing mistrust on the government and mainstream media because the radiation records release was delayed (Tsang, 2013).

One of the solid crowd initiations which was facilitated by SNSs was citizens' participation in measuring radiation caused by the incident. Citizens started buying Geiger counters, running out commercially available supplies instantly. At this point the discussion turned to building one's own devices. This discussion was taken up in the Tokyo Hackerspace by a multidisciplinary group of people and as a result in a week "bGeigie" was built and that was just the beginning of the Safecast initiative. Safecast (blog.safecast.org) empowered citizens to build their own Geiger devices and carry them around to collect radiation measures. This enabled people to easily monitor their own environments, and to not depend on governmental bodies for this kind of vital information. All the collected data is open so anyone can use it. Within short time this initiative attracted world experts and became the go-to independent source of information on radiation issues around the world. Safecast was the result of utilizing the potential in social network services in a chaotic situation. By now, Safecast has been able to collect more radiation data than all projects in history and is biggest monitoring project that has ever existed. Safecast is one example of SNS data utilization for intelligent insight - often also referred as citizen science, crowd-sensing, and crowdsourcing – and the data being submitted in this sort of system represents a deliberative act of public participation by the interaction of the public with technology-enabled services. The evolutionary pattern of discussion in SNSs apparently resulted in a better understanding of

the problems which resulted in applications such as Safecast. Over the last decade, similar crowd participatory initiatives led to applications such as eBird, Fold.it, Waze, Ushahidi, and Galaxyzoo. These applications have been actively supporting citizen-driven data collection for a variety of purposes including scientific research and crisis communication, whilst serving as means for inclusive engagement, education, and public outreach. Regarding the Fukushima incident, we will investigate how the characteristics of the major stakeholders and content producers have been changed over the time. Users in Twitter create content while resharing posts by retweeting them. In our case we merged the retweet posts as the intention is to understand how the original content in Twitter happen to evolve from a topical prospective. By using the Twitter API, we constructed our searching query by including major hashtags around the topic in both English and Japanese from the starting day of the incident (March 11th, 2011) to the day of data collection (January 11th, 2017). We used the implemented automatic language detection system of Twitter to identify the language of tweets (latter the local Japanese tweets were translated to be included for the content topical analysis). The Twitter analytic process was facilitated by Azure cloud computing platform (azure.microsoft.com) and the process will be elaborated in the Method section.

Method

In this paper, we present the overall architecture of the procedure which has been applied to get insight from SNS data. We consider data collected on Twitter (twitter.com), a microblogging platform used by millions of bloggers. On Twitter, each user can freely post

short messages (up to 140 characters) called “tweets” to their followers. Twitter provides application programming interfaces (APIs) to access tweets and information about tweets and users. The potential bias of Twitter APIs was discussed by recent research, such the access to the random sample set of tweets rather than the full data (González-Bailón, Wang, Rivero, Borge-Holthoefer, & Moreno, 2014). In order to retrieve relevant data regarding the Fukushima incident, a searching query has been constructed for the Twitter API. In the searching query, various specifications can be implemented such as keywords, length, and date to target the topic of interest. In our case we construct a query out of a set of keywords (“fukushima”, “radioactive”, “nuclear power”, “reactors”, “radioactive”, “meltdown”, “radiation”, “earthquake”, “power plant”). Within the specified time frame regarding the case study (March 11th, 2011 till January 11th, 2017), 163,000 tweets were retrieved. A systematic process has designed and utilized to get the required insight from the collected SNS data which is composed and presented graphically in Figure 2.

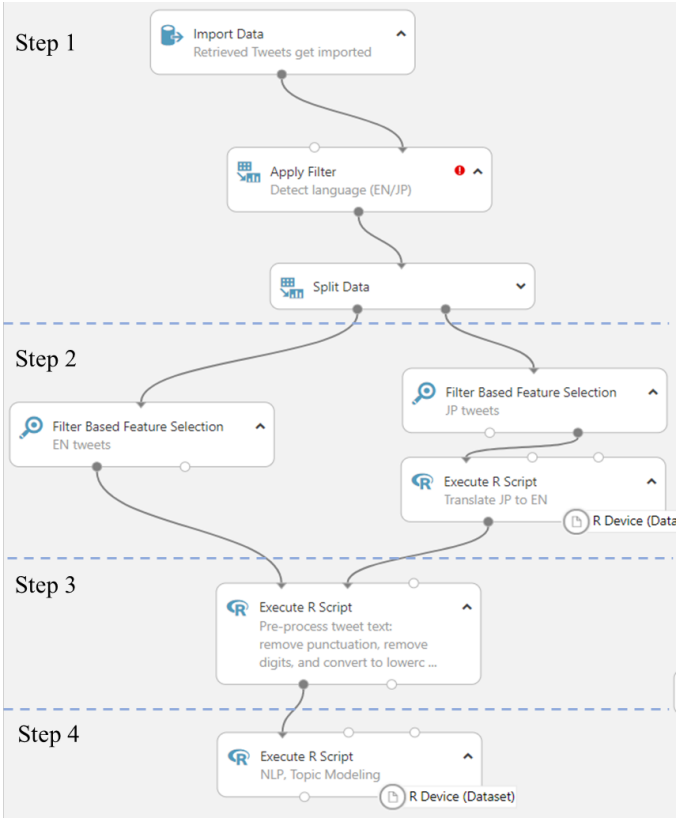


Figure 2. Twitter content analysis with Azure Cloud Computing Platform

Step 1: Data retrieval methods are often loosely controlled, resulting in out of range values. The data pre-processing task is performed to reduce the irrelevant and redundant data present in the collected set. The task includes importing data retrieved from SNSs (in our case Twitter) and applying the range of filters to first detect the language (English or Japanese) and thereafter split the data based on the language. This task is

necessary for the forthcoming steps so to normalize the data for a better knowledge discovery results.

Step 2: This task will derive values (features) from the data regarding the context of the knowledge discovery process. The package 'translateR' from R programming language has been used to convert the Japanese tweets to the equivalent English translations. The intent for feature extraction is to facilitate the further distinctions and categorization of the data. In this process, stopwords and stemming have also been utilized on the combined bag of tweets (English and translated Japanese tweets) for better preparation for the next analytical tasks.

Step 3: Classification of the data occurs in order to reduce the dimensionality of the data. It is an approach derived from the general hypothesis of the knowledge discovery task so to distinguish the best fit data points from the mass. In our case study, topic modeling has been performed so to understand the evolution of discussions on Twitter regarding the Fukushima incident over the time. Topic modelling can be described as a method for finding a group of words (i.e. a topic) from a collection of documents that best represents the information in the collection. It can also be thought of as a form of text mining – a way to obtain recurring patterns of words in textual material (Sievert & Shirley, 2014). The separate script to perform the topic modeling calculation is batched with the Microsoft azure cloud (Microsoft, 2017).

Step 4: The insights from the results can be provided in a visually appealing way which will be explored further in the next section.

Results and Summery

The main reason to apply the unsupervised learning methods with the large amount of tweets is to reduce the dimensionality of the retrieved text for better insight. In general, we are interested in learning about the latent similarities with discussion topics in three periods of the study and the relative collected data. Topic modeling uses a nonparametric Bayesian model to measure similarity between documents and measure topics. There are many techniques that are used to obtain topic models; in this study, we leveraged Latent Dirichlet Allocation (LDA). We use LDA as a discriminative model for classifying tweets. The validation accuracy was maximized when there were 15 LDA topics. In other words, the model is the best explanatory of discussion topic distribution when the clustering is set for 15 topics. The motivation in our analysis is to observe the distribution of topics of discussion in SNS and examine the topical proximity distance as the time evolves. Regarding the analysis of tweets collected for the Fukushima incident, after data retrieval, processing and topical analysis, the matrices presented in Figure 3 is one application to show the probabilistic distribution of generated discussions topics explain the most variance in the data.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Topic 1	100	19	15	44	60	8	83	97	20	61	61	53	9	70	61
Topic 2	66	100	87	100	39	22	16	45	44	48	70	68	19	37	69
Topic 3	24	46	100	38	8	34	37	61	82	84	95	26	53	38	5
Topic 4	9	87	46	100	17	68	94	45	21	30	67	11	5	62	46
Topic 5	11	82	86	33	100	7	40	87	0	47	24	85	87	11	29
Topic 6	4	37	43	69	64	100	50	34	90	30	64	24	24	44	84
Topic 7	83	48	2	57	92	34	100	84	80	97	93	48	75	99	39
Topic 8	53	49	87	1	52	50	23	100	53	67	65	11	83	51	55
Topic 9	69	16	81	48	59	42	100	47	100	53	34	43	14	68	14
Topic 10	44	52	97	91	41	74	79	26	26	100	48	64	64	20	82
Topic 11	14	26	86	68	67	71	25	99	7	94	100	47	88	31	33
Topic 12	17	84	61	39	36	56	11	79	71	1	99	100	5	27	5
Topic 13	41	27	86	10	55	17	79	26	100	93	50	75	100	46	22
Topic 14	97	20	46	18	66	33	61	0	59	48	13	21	28	100	75
Topic 15	89	23	87	88	16	0	25	40	23	76	71	84	69	4	100

Period 1: 2011-2013

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Topic 1	100	60	68	50	71	39	63	9	40	67	47	33	67	19	33
Topic 2	78	100	67	41	56	43	10	63	1	19	16	5	39	5	35
Topic 3	79	94	100	65	52	36	17	44	7	13	19	20	6	0	58
Topic 4	79	76	83	100	74	77	72	24	14	47	51	66	10	38	11
Topic 5	9	71	42	83	100	66	38	8	40	65	63	75	30	41	18
Topic 6	68	20	8	75	75	100	44	62	68	32	64	75	34	1	26
Topic 7	25	41	63	75	93	84	100	85	100	90	62	64	39	34	13
Topic 8	65	55	3	64	34	5	87	100	87	86	79	4	16	39	7
Topic 9	26	48	67	2	11	61	81	75	100	86	63	96	93	27	55
Topic 10	42	37	21	65	42	5	62	77	85	100	75	82	95	12	23
Topic 11	56	27	3	68	67	47	92	61	70	8	100	76	84	86	74
Topic 12	4	60	24	72	61	48	58	10	49	68	69	100	75	67	86
Topic 13	52	20	73	33	61	64	67	3	35	50	53	29	100	96	85
Topic 14	52	3	38	18	64	28	57	39	18	47	1	67	19	100	79
Topic 15	50	73	3	28	68	47	38	70	48	50	1	22	39	66	100

Period 2: 2013-2015

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Topic 1	100	95	97	83	71	39	63	9	40	67	47	53	67	19	33
Topic 2	80	100	89	98	56	43	10	63	1	19	16	5	39	5	35
Topic 3	75	87	100	94	52	36	17	44	7	13	19	20	6	0	58
Topic 4	83	95	79	100	94	80	72	24	14	47	51	66	10	38	11
Topic 5	9	71	42	77	100	79	38	8	40	65	63	75	30	41	18
Topic 6	68	20	8	81	78	100	93	62	68	32	64	75	34	1	26
Topic 7	25	41	63	98	82	96	100	90	79	81	81	64	39	34	13
Topic 8	2	33	25	35	8	5	78	100	94	99	87	4	16	39	7
Topic 9	23	33	39	24	15	61	66	79	100	92	77	77	60	27	55
Topic 10	22	4	25	30	1	5	87	67	97	100	92	87	77	12	23
Topic 11	24	15	24	20	9	47	94	80	70	8	100	81	75	90	64
Topic 12	30	0	27	30	35	13	20	19	36	68	81	100	97	93	87
Topic 13	39	8	29	16	15	20	11	31	13	10	28	29	100	90	79
Topic 14	17	39	36	6	39	18	2	12	32	28	11	67	97	100	99
Topic 15	39	36	19	11	24	33	8	22	31	24	38	22	87	98	100

Period 3: 2015-2017

0
10
20
30
40
50
60
70
80
90
100

Estimated distance proximity range

Figure 3. Estimated average distance between discrete topic discussions over three time periods

Topic models are generative which means that they model texts as if they were generated from a certain probability distribution. In our case, each tweet defines a distribution over (hidden) topics or a distribution over words. The posterior probability of latent variables given a corpus determines the collection of tweets into topics. Each topic represents a bag of tweets which contains the ones with closer word pattern usage similarity. In other words, similar tweets are clustered among each other. The trend has been plotted with three separated

time periods (2011-2013, 2013-2015, and 2015-2017) so to enable the comparison of topical evolution on SNS discussions. Our analysis shows that the topical structure of the discussion progresses to a closer proximity in period three from a more topical distance in beginning of the discussion in period one. It is clearly visible that the discussions topical coordination have merged closer to a certain diagonal position in the matrix in the third period while in the first period the 15 detected topical discussion are positioned with a lower estimated distance from each other.

Discussion and Concluding Remarks

We have noted that the emergence of digital technologies in today's world introduces complexity where it is difficult to gain a deeper scholarly understanding of the structure, dynamics, and strategy/behavior of platforms and associated components in the ecosystems around digital platforms. As a result of the widespread diffusion of large-scale data, there is loss of direct executive control as the complexity across services escalates. The emerging data sources can provide important insight to understand digital platforms and ecosystem information for different levels and scopes of analysis. A huge amount of data on collective behaviors is being generated from SNSs which permeate all levels of society. This phenomenon promotes quantitative analysis of these data, with the goal to understand collective behaviors and predict them in effective and efficient ways. SNSs are a necessary component in the digital ecosystem and should be leveraged in their full capacity. In this paper, we discuss the fundamentals of a digital ecosystem and the role of SNSs as the

emerging component representing society and the embedded social capital. The ability to derive insightful information from digital platforms can be facilitated by recent advancements in text analysis and low-cost accessibility to powerful cloud-based toolkits. We proposed a systematic process to analyze SNS data to leverage the full potential and accurate insight. To put the systematic process into practice, we analyzed dynamics of discussions during a disaster, specifically the Fukushima nuclear power plant crisis. In particular, we have carried out textual content analysis periodically, to obtain the topical evolution of discussion in SNS (in our case Twitter) regarding the Fukushima incident.

The motivation for this experiment was to understand the role of SNSs as a major component in digital platforms for the matters of information flow and community discussion evolution. Moreover, advancements in data retrieval, textual analysis, and topic modeling have been employed for deriving information from the massive SNS data. We found that proximity topic of discussion gets closer over time which forms cohesion. Due to our literature study on the incident we identify this cohesion of discussion to be related to various reasons such as:

- Appearance of major stakeholders on Twitter which guided the discussions (i.e. Safecast).
- Distinction of active and reputable content generator profiles over time.
- The process of having explicit problems and needs dominated in discussion over time.

- Societal education over the subject matured and progressed over time with certain direction.

Social Network Services in a digital ecosystem is unique to the context and object of the study. Understanding the newly emerging SNSs can help to define the footprint or scaffolding for making a progressive, dynamic framework that will steer towards achieving digital objectives. The development in computational power and advancement in big data analytics requires systematic approaches for leveraging SNS data for valuable insights. This study introduces a systemic approach toward analysis of SNS data. This agenda further defines the advancements in computational capabilities necessary for developing accurate competencies based on SNS textual content.

Finally, we address a few shortcomings of our study and discuss how to improve it with extended analysis. In this study, the influence of highly reputable profiles and the quality of the content which they are generating has not been considered, although in reality users will typically consider the profile and the quality of content before interacting with it. Presumably, the direction for interaction with information on Twitter correlates with the quality of the information generator and the content itself. In this direction, our analysis may be extended to take into account the wording of tweets; probing other motivations to tweet or retweet can also refine the model. These issues are left for future study.

REFERENCES

Adner, R. (2017). Ecosystem as Structure: An Actionable Construct for Strategy. *Journal of Management*, *43*(1), 39–58. http://doi.org/10.1177/0149206316678451

Albino, V., Berardi, U., & Dangelico, R. M. (2015). Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, *22*(1), 1–19. http://doi.org/10.1080/10630732.2014.942092

Anderson, M., & Caumont, A. (2014). How social media is reshaping news. Retrieved September 6, 2017, from http://www.pewresearch.org/fact-tank/2014/09/24/how-social-media-is-reshaping-news/

Arthur, W. B. (1989). Competing Technologies, Increasing Returns, and Lock-In by Historical Events. *The Economic Journal*, *99*(394), 116. http://doi.org/10.2307/2234208

Brunswicker, S., Matei, S. A., Zentner, M., Zentner, L., & Klimeck, G. (2017). Creating impact in the digital space: digital practice dependency in communities of digital scientific innovations. *Scientometrics*, *110*(1), 417–442. http://doi.org/10.1007/s11192-016-2106-z

Clark, K. B. (1985). The interaction of design hierarchies and market concepts in technological evolution. *Research Policy*, *14*(5), 235–251. http://doi.org/10.1016/0048-7333(85)90007-1

Cooper, G. P., Yeager, V., Burkle, F. M., & Subbarao, I. (2015). Twitter as a potential disaster risk reduction tool. part i: Introduction, terminology, research and operational applications. *PLoS Currents*, *7*(DISASTERS). http://doi.org/10.1371/currents.dis.a7657429d6f25f02bb5253e551015f0f

de Reuver, M., Sørensen, C., & Basole, R. C. (2017). The digital platform: a research agenda. *Journal of Information Technology*. http://doi.org/10.1057/s41265-016-0033-3

Dew, N., & Read, S. (2007). The more we get together: Coordinating network externality product introduction in the RFID industry. *Technovation*, *27*(10), 569–581. http://doi.org/doi.org/10.1016/j.technovation.2006.12.005

Eom, Y. H., Puliga, M., Smailovic, J., Mozetic, I., & Caldarelli, G. (2015). Twitter-based analysis of the dynamics of collective attention to political parties. *PLoS ONE*, *10*(7). http://doi.org/10.1371/journal.pone.0131184

European commision. (2016). *Accelerating the digital transformation of European industry and enterprises*. *Strategic Policy Forum on Digital Entrepreneurship*. Retrieved from

https://www.google.fi/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=0ahUKEwinvfGfkebWAhWLFJoKHUuVBIYQFgg0MAM&url=http%3A%2F%2Fec.europa.eu%2FDocsRoom%2Fdocuments%2F15856%2Fattachments%2F1%2Ftranslations%2Fen%2Frenditions%2Fnative&usg=AOvVaw2aCnE5hH2Kz1wVnZ

Eurostat. (2017). Digital economy and society statistics - households and individuals. Retrieved September 5, 2017, from http://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals

Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, *43*(7), 1239–1249. http://doi.org/10.1016/j.respol.2014.03.006

Gawer, A., & Cusumano, M. A. (2014). Industry platforms and ecosystem innovation. *Journal of Product Innovation Management*, *31*(3), 417–433. http://doi.org/10.1111/jpim.12105

González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y. (2014). Assessing the bias in samples of large online networks. *Social Networks*, *38*(1), 16–27. http://doi.org/10.1016/j.socnet.2014.01.004

Hajikhani, A. (2018). University-Industry Programs as Platforms: A Case Study of MultiDisciplinary Collaborative Network Development. In *51st Hawaii International Conference on System Sciences*.

Henfridsson, O., Mathiassen, L., & Svahn, F. (2014). Managing technological change in the digital age: The role of architectural frames. *Journal of Information Technology*, *29*(1), 27–43. http://doi.org/10.1057/jit.2013.30

Hirschburg, P. (1986). Media system dependency theory: Responses to the eruption of Mount St. Helens. *Media, Audience, and* …. Retrieved from https://scholar.google.com/scholar?q=media+system+dependency+theory&btnG=&hl=en&as_sdt=0%2C7#4

Hughes, A. L., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, *6*(3/4), 248. http://doi.org/10.1504/IJEM.2009.031564

Jaeger, P. T., Shneiderman, B., Fleischmann, K. R., Preece, J., Qu, Y., & Fei Wu, P. (2007). Community response grids: E-government, social networks, and effective emergency management. *Telecommunications Policy*, *31*(10–11), 592–604. http://doi.org/10.1016/j.telpol.2007.07.008

Jung, J. Y. (2012). Social Media Use and Goals after the Great East Japan Earthquake.

Kallinikos, J., Aaltonen, A., & Marton, A. (2013). The Ambivalent Ontology of Digital Artifacts. *MIS Quarterly*, *37*(2), 357–370. http://doi.org/10.25300/MISQ/2013/37.2.02

Katz, M. L., & Shapiro, C. (1985). Network Externalities, Companition and Compability. *American Economic Review, Vol. 75*. Retrieved from http://brousseau.info/pdf/cours/Katz-Shapiro[1985].pdf

Kim, Y. C., Jung, J. Y., Cohen, E. L., & Ball-Rokeach, S. J. (2004). Internet connectedness before and after September 11 2001. *New Media and Society*, *6*(5), 611–631. http://doi.org/10.1177/146144804047083

Kireyev, K., Palen, L., & Anderson, K. (2009). Applications of topics models to analysis of disaster-related twitter data. *NIPS Workshop on Applications for Topic Models: Text and Beyond*. Retrieved from http://www.umiacs.umd.edu/~jbg/nips_tm_workshop/15.pdf

Legner, C., Eymann, T., Hess, T., Matt, C., Böhmann, T., Drews, P., … Ahlemann, F. (2017). Digitalization: Opportunity and Challenge for the Business and Information Systems Engineering Community. *Business & Information Systems Engineering*, *59*(4), 301–308. http://doi.org/10.1007/s12599-017-0484-2

Lindsay, B. R. (2011). Social Media and Disasters: Current Uses, Future Options and Policy Considerations. *Congressional Research Service Reports*, 13. Retrieved from http://fas.org/sgp/crs/homesec/R41987.pdf

Lowrey, W. (2004). Media Dependency During a Large-Scale Social Disruption: The Case of September 11. *Mass Communication and Society*, *7*(3), 339–357. http://doi.org/10.1207/s15327825mcs0703_5

Meadows, D. (2011). *Thinking in Systems: A Primer*. *Environmental Politics* (Vol. 20). http://doi.org/10.1080/09644016.2011.589585

Microsoft. (2017). Execute R Script in Microsoft Azure. Retrieved from https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/execute-r-script

Muralidharan, S., Rasmussen, L., Patterson, D., & Shin, J. H. (2011). Hope for Haiti: An analysis of Facebook and Twitter usage during the earthquake relief efforts. *Public Relations Review*, *37*(2), 175–177. http://doi.org/10.1016/j.pubrev.2011.01.010

OECD. (2016). Digital Government Strategies for Transforming Public Services in the Welfare Areas, 1–63. Retrieved from http://www.oecd.org/gov/digital-government/Digital-Government-Strategies-Welfare-Service.pdf

Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational Collaboration

and the Locus of Innovation: Networks of Learning in Locus of Innovation: Networks of Learning in Biotechnology. *Source: Administrative Science Quarterly*, *41*(1), 116–145. http://doi.org/10.2307/2393988

Shapiro, C., & Varian, H. R. (1998). Information rules: a strategic guide to the network economy.

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70. Retrieved from http://www.aclweb.org/anthology/W/W14/W14-3110

Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010). Chatter on the red: what hazards threat reveals about the social life of microblogged information. *CSCW '10 Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, 241–250. http://doi.org/10.1145/1718918.1718965

Tilson, D., Lyytinen, K., & Sørensen, C. (2010). Digital infrastructures: The Missing 15 Research Agenda. *Information Systems Research*, *21*(4), 748–759. http://doi.org/10.1287/isre.ll00.0318

Tilson, D., Sørensen, C., & Lyytinen, K. (2012). Change and control paradoxes in mobile infrastructure innovation: The Android and iOS mobile operating systems cases. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 1324–1333. http://doi.org/10.1109/HICSS.2012.149

Tsang, M. (2013). Lessons from Fukushima: Do not ignore citizen media. Retrieved from https://www.worldenergy.org/news-and-media/news/lessons-from-fukushima-do-not-ignore-citizen-media

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*, 1079. http://doi.org/10.1145/1753326.1753486

World Economic Forum. (2017). Mapping Global Transformations. Retrieved from https://www.weforum.org/about/transformation-maps

Yoo, Y., & Henfridsson, O. (2010). The New Organizing Logics of Digital Innovation : An Agenda for Information Systems Research The New Organizing Logics of Digital Innovation : An Agenda for Information Systems Research. *Information Systems Research*, (1), 1–20. http://doi.org/10.1287/isre.1100.0322

ACTA UNIVERSITATIS LAPPEENRANTAENSIS

773. KURAMA, ONESFOLE. Similarity based classification methods with different aggregation operators. 2017. Diss.

774. LYYTIKÄINEN, KATJA. Removal of xylan from birch kraft pulps and the effect of its removal on fiber properties, colloidal interactions and retention in papermaking. 2017. Diss.

775. GAFUROV, SALIMZHAN. Theoretical and experimental analysis of dynamic loading of a two-stage aircraft engine fuel pump and methods for its decreasing. 2017. Diss.

776. KULESHOV, DMITRII. Modelling the operation of short-term electricity market in Russia. 2017. Diss.

777. SAARI, JUSSI. Improving the effectiveness and profitability of thermal conversion of biomass. 2017. Diss.

778. ZHAO, FEIPING. Cross-linked chitosan and β-cyclodextrin as functional adsorbents in water treatment. 2017. Diss.

779. KORHONEN, ILKKA. Mobile sensor for measurements inside combustion chamber – preliminary study. 2017. Diss.

780. SIKIÖ, PÄIVI. Dynamical tree models for high Reynolds number turbulence applied in fluid-solid systems of 1D-space and time. 2017. Diss.

781. ROMANENKO, ALEKSEI. Study of inverter-induced bearing damage monitoring in variable-speed-driven motor systems. 2017. Diss.

782. SIPILÄ, JENNI. The many faces of ambivalence in the decision-making process. 2017. Diss.

783. HAN, MEI. Hydrodynamics and mass transfer in airlift bioreactors; experimental and numerical simulation analysis. 2017. Diss.

784. ESCALANTE, JOHN BRUZZO. Dynamic simulation of cross-country skiing. 2017. Diss.

785. NOKKA, JARKKO. Energy efficiency analyses of hybrid non-road mobile machinery by real-time virtual prototyping. 2018. Diss.

786. VUORIO, ANNA. Opportunity-specific entrepreneurial intentions in sustainable entrepreneurship. 2018. Diss.

787. PULKKINEN, AKI. Towards a better understanding of activity and selectivity trends involving K and O adsorption on selected metal surfaces. 2017. Diss.

788. ZHAO, WENLONG. Reliability based research on design, analysis and control of the remote handling maintenance system for fusion reactor. 2018. Diss.

789. IAKOVLEVA, EVGENIA. Novel sorbents from low-cost materials for water treatment. 2018. Diss.

790. KEDZIORA, DAMIAN. Service offshoring industry: systems engineering approach to its transitional challenges. 2018. Diss.

791. WU, JING. Soft computing methods for performance improvement of EAMA robot in fusion reactor application. 2018. Diss.

792. VOSTATEK, PAVEL. Blood vessel segmentation in the analysis of retinal and diaphragm images. 2018. Diss.

793. AJO, PETRI. Hydroxyl radical behavior in water treatment with gas-phase pulsed corona discharge. 2018. Diss.

794. BANAEIANJAHROMI, NEGIN. On the role of enterprise architecture in enterprise integration. 2018. Diss.

795. HASHEELA-MUFETI, VICTORIA TULIVAYE. Empirical studies on the adoption and implementation of ERP in SMEs in developing countries. 2018. Diss.

796. JANHUNEN, SARI. Determinants of the local acceptability of wind power in Finland. 2018. Diss.

797. TEPLOV, ROMAN. A holistic approach to measuring open innovation: contribution to theory development. 2018. Diss.

798. ALBATS, EKATERINA. Facilitating university-industry collaboration with a multi-level stakeholder perspective. 2018. Diss.

799. TURA, NINA. Value creation for sustainability-oriented innovations: challenges and supporting methods. 2018. Diss.

800. TALIKKA, MARJA. Recognizing required changes to higher education engineering programs' information literacy education as a consequence of research problems becoming more complex. 2018. Diss.

801. MATTSSON, ALEKSI. Design of customer-end converter systems for low voltage DC distribution from a life cycle cost perspective. 2018. Diss.

802. JÄRVI, HENNA. Customer engagement, a friend or a foe? Investigating the relationship between customer engagement and value co-destruction. 2018. Diss.

803. DABROWSKA, JUSTYNA. Organizing for open innovation: adding the human element. 2018. Diss.

804. TIAINEN, JONNA. Losses in low-Reynolds-number centrifugal compressors. 2018. Diss.

805. GYASI, EMMANUEL AFRANE. On adaptive intelligent welding: Technique feasibility in weld quality assurance for advanced steels. 2018. Diss.

806. PROSKURINA, SVETLANA. International trade in biomass for energy production: The local and global context. 2018. Diss.

807. DABIRI, MUHAMMAD. The low-cycle fatigue of S960 MC direct-quenched high-strenght steel. 2018. Diss.

808. KOSKELA, VIRPI. Tapping experiences of presence to connect people and organizational creativity. 2018. Diss.

809. HERALA, ANTTI. Benefits from Open Data: barriers to supply and demand of Open Data in private organizations. 2018. Diss.

810. KÄYHKÖ, JORMA. Erityisen tuen toimintaprosessien nykytila ja kehittäminen suomalaisessa oppisopimuskoulutuksessa. 2018. Diss.

Acta Universitatis
Lappeenrantaensis
811

ISBN 978-952-335-265-0
ISBN 978-952-335-266-7 (PDF)
ISSN-L 1456-4491
ISSN 1456-4491
Lappeenranta 2018