

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/125585>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Three Essays on Growth and Innovation of Digital Platforms

A thesis submitted in complete fulfilment of the
requirements for the degree Doctor of Philosophy

Submitted by:

Philipp Hukal

Submitted to:

Information Systems & Management Group
Warwick Business School – The University of Warwick

August 2018

Table of Contents

CHAPTER 1 – INTRODUCTION	1
1.1. DIGITAL INNOVATION	3
1.1.1. Definition.....	4
1.1.2. Digital Technology.....	4
1.1.3. Recombination.....	5
1.1.4. Co-Creation	7
1.2. PLATFORMS AS MEANS OF ORGANISING DIGITAL INNOVATION	8
1.3. EMPIRICAL SETTING	10
1.4. STRUCTURE OF THE DISSERTATION	13
CHAPTER 2 – LITERATURE REVIEW	14
2.1. MAJOR STREAMS IN RESEARCH ON DIGITAL PLATFORMS	14
2.1.1. Platforms as Marketplaces.....	15
2.1.2. Platforms as Modular Systems	18
2.1.3. Platforms as Ecosystems	22
2.1.4. Platforms as Socio-Technical Systems.....	25
2.2. POTENTIAL CONTRIBUTIONS	26
2.2.1. Detailed Empirical Accounts.....	27
2.2.2. Generativity	27
2.2.3. Design and Organisation of Digital Technology.....	29
CHAPTER 3 – ANALYTICAL APPROACH.....	31
3.1. COMPUTATIONAL RESEARCH IN IS	31
3.1.1. The core activities in computational IS research.....	34
3.1.2. Digital process trace data	35
3.1.3. The IS artifact.....	37
3.2. MIXED-METHODS DESIGNS	39
3.2.1. Pragmatist Influence.....	40
CHAPTER 4 – THE IMPACT OF ENDORSEMENTS ON PLATFORM GENERATIVITY	46
ABSTRACT	46
4.1. INTRODUCTION	47
4.2. PRIOR LITERATURE AND CONCEPTUAL BASIS	48
4.2.1. Endorsements and Generativity.....	50
4.3. RESEARCH DESIGN	52
4.3.1. Empirical Context.....	52
4.3.2. The role of metadata for generativity	54
4.3.3. Analytical Approach.....	55
4.4. ANALYSIS STAGE 1: CONTENT ANALYSIS AND HYPOTHESES	56
4.4.1. Data	57
4.4.2. Endorsement Motives.....	59
4.5. ANALYSIS STAGE 2: CONFIRMATORY HYPOTHESIS TESTS	67
4.5.1. Data	68
4.5.2. Endorsement and Platform Scale	70
4.5.3. Endorsement and Platform Scope	71
4.6. RESULTS	73
4.6.1. Robustness Checks	78
4.7. DISCUSSION	79
4.7.1. Successful signals of endorsement	80
4.7.2. Unsuccessful signals of endorsement.....	82
4.8. IMPLICATIONS	84
4.9. CONCLUSION.....	86
APPENDIX TO CHAPTER 4.....	87

CHAPTER 5 – COMPLEXITY, EXTERNAL DEPENDENCE, AND THE PACE OF CHANGE IN UNBOUNDED DIGITAL PLATFORM ECOSYSTEMS	91
ABSTRACT	91
5.1. INTRODUCTION	92
5.2. CONCEPTUALISING UNBOUNDED PLATFORM ECOSYSTEMS	94
5.2.1. Unboundedness of Digital Platform Ecosystems	94
5.2.2. Modules in Unboundedness Ecosystems	96
5.2.3. Organising Unbounded Platform Ecosystems.....	98
5.3. THEORY DEVELOPMENT AND HYPOTHESES	100
5.4. EMPIRICAL ANALYSIS	102
5.4.1. Context	102
5.4.2. Data	103
5.4.3. Analytical Approach.....	103
5.4.4. Measures.....	104
5.4.5. Hypothesis Tests.....	110
5.6. RESULTS	111
5.6.1. Cross-Module Analysis – Comparison.....	111
5.6.2. Within-Module Analysis – Panel Regressions.....	114
5.6.3. Robustness Checks	119
5.7. DISCUSSION	120
APPENDIX TO CHAPTER 5.....	122
CHAPTER 6 – PERICENTRIC COORDINATION ON DIGITAL PLATFORMS DURING BOUNDARY RESOURCE INSTABILITY	127
6.1. INTRODUCTION	128
6.2. CONCEPTUAL BACKGROUND	130
6.2.1. Platform Boundary Resources.....	131
6.2.2. Participant Coordination on Digital Platforms.....	132
6.3. RESEARCH DESIGN	135
6.3.1. Empirical Setting.....	135
6.3.2. Data	136
6.3.3. Analysis and Framing.....	139
6.4. FINDINGS	143
6.4.1. Deviation from Core-Periphery Coordination.....	143
6.4.2. Coordination among participants	146
6.5. THREATS TO VALIDITY	152
6.6. DISCUSSION	155
6.5.1. Pericentric Coordination.....	155
6.7. CONCLUSION.....	160
APPENDIX TO CHAPTER 6.....	163
ADDENDUM: PRE-STUDY	165
CHAPTER 7 – SUMMARY	173
REFERENCES.....	176

List of Tables

Table 2.1. Major Streams of Research on Digital Platforms	15
Table 3.1. Analytical Approach in the Three Empirical Studies	44
Table 4.1. Identified Motives for Endorsement	66
Table 4.2. Variable Definitions	69
Table 4.3. Model Results – Platform Scale (Logistic Regression)	75
Table 4.4. Model Results – Platform Scope (Quasi-Poisson Regression)	77
Table 4.5. Overview of Hypothesis Test Results	77
Table 4.6.A. Full Results: Chi-Square Test of Independence	87
Table 4.7.A. Descriptive Statistics (Dataset 1)	90
Table 4.8.A. Descriptive Statistics (Dataset 2)	90
Table 5.1. Descriptive Statistics	108
Table 5.2. Variable Definitions	109
Table 5.3. Mean Values per Design Pattern	111
Table 5.4. Model Results – Mixed Effects Panel Regressions	118
Table 5.5.A. Module Overview	124
Table 5.6.A. Queries used in Source Code Analysis	126
Table 6.1. Data Collection	138
Table 6.2. Overview of E-Mail Discussion Characteristics	151
Table 6.3.A. Keywords Used in Data Filtering	163
Table 6.4.A. Downloaded E-Mail Lists by Topic	163
Table 6.5.A. List of Informant Interviews (primary)	164
Table 6.6.A. OLS Model Results	164

List of Figures

Figure 1.1. Interest in Digital Platforms	2
Figure 4.1. OSM Tag Combinations	53
Figure 4.2. Tag Edits on an OSM highway Object over Time	54
Figure 4.3. Illustrated Results from Logistic Regression	74
Figure 4.4. Illustrated Results from Quasi-Poisson Regression	76
Figure 4.5.A. Histogram: Tag Co-Occurrence	88
Figure 4.6.A. Mean-Variance Relationship in Sample	88
Figure 5.1. Visualised Hypotheses in 2x2 Framework	102
Figure 5.2. Comparison Results in 2x2 Framework	113
Figure 5.3. Model Results in 2x2 Framework	116
Figure 6.1. Timeline Visualisation	143
Figure 6.2. Relationship between Communication and Contributions	144
Figure 6.3. Levels of Communication Intensity	145
Figure 6.4. Measures of Productivity by Level of Communication Intensity	146
Figure 6.5. Visualisation of Participant Coordination and Contribution	154
Figure 6.6. Schematic Representation of Pericentric Coordination	159

Acknowledgements

This dissertation would not have been possible without immense support of many colleagues, friends, and family.

First and foremost, I thank my two outstanding supervisors, Professor Ola Henfridsson, and Professor Maha Shaikh at Warwick Business School. It was through their constant help, advice, and guidance that I learned to communicate ideas to an academic audience. I am grateful that they believed in me and went out of their way to see my – and later our joint work – succeed. I remain indebted to their supervision and should I ever amount to a half-decent IS researcher, it is because of these two. Where you found the patience to deal with me, continues to puzzle me.

Over the years, I have had the opportunity to spend time with truly great thinkers in our field, some of which have become important advisors along the way. I would like to particularly thank Professor Youngjin Yoo and Professor Nicholas Berente for opening their offices and homes to me and taking the time to sit down and discuss ideas, give pointers on the craft of researching, and for their kindness of sharing their visions. I am thankful for all your support and look forward to continuing our collaborations.

The PhD Programme at WBS comprises an impressive group of people that became true friends and sparring partners in all things PhD. Sharing an office with Danilo Cascaldi Garcia and Krishane Patel was one of the best things that could have happened to me. Kolja Johanssen remains a close friend, as does Karsten Müller from whom I have learned a lot. Iram Ahmad's walks ensured my sanity at times. My thanks go to all of them and many others in the 2014 cohort.

I also have the great fortune of being supported by three extraordinarily strong women. First, my grandmother who never had her doctoral degree conferred. Her strength and tenacity remain a source of inspiration. My mother, who endured hours of nerdy lectures on the subject of this research, and yet never tired of sharing wisdom and love. Standing in the lineage of these two fills me with pride.

Finally, my gratitude goes to my partner Celia who has supported me throughout the PhD. Her understanding, encouragement, and listening, not to mention proof-reading

and editing, helped me get through this journey. The Sam to my Frodo-like existence, it is thanks to her that I did not carry the ring alone.

This dissertation would have been impossible without every single one of you. These few words alone cannot express how truly and deeply grateful I am for everything you have done.

London / Berlin / Copenhagen

Summer 2018

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented – including data generated and data analysis – was carried out by myself. All mistakes remain my own.

Parts of this thesis have been submitted for publication:

Chapter 1 – Digital Innovation

Chapter 4 – The Impact of Endorsements on Platform Generativity

*Chapter 5 – Complexity, External Dependence, and the Pace of Change in
Unbounded Platform Ecosystems*

Abstract

Digital platforms are complex digital technology arrangements that enable the interaction of otherwise unaffiliated organisations. This interaction often generates novel outputs and as a result digital platforms are seen as a powerful driver of digital innovation. Yet exactly how digital platforms generate innovations by facilitating interaction merits further investigation. This dissertation illustrates aspects of how platforms grow and innovate using the case of the open-geo data platform OpenStreetMap. The study draws from both quantitative as well as qualitative analysis techniques applied to highly detailed data capturing the use, design, and operation of the platform over more than ten years. A series of computationally-intensive, mixed-methods studies were conducted to utilise the full scale of available empirical material while maintaining contextual richness relevant to the case. Embedded in recent topics on digital platforms, three empirical studies are presented. Each study focuses on one aspect of growth and innovation on digital platforms. The studies specifically examine; (i) how platform operators can stimulate generativity, that is the generation of novel outputs without direct input by the operator, (ii), how the unique attributes of digital technologies enable the creation of complex ecosystems that allow for high-paced changes in a platform's architecture even if that increases the structural complexity of a platform, and, (iii) how participants coordinate contributions to a platform's operation when they cannot rely on stable interfaces. Collectively these studies contribute to the understanding of how platforms generate new digital innovations.

Abbreviations

API – Application Programming Interface

CF – Control Function

GLM – General Linear Modelling

GPS – Global Positioning System

IV – Instrumental Variable

IRC – Internet Relay Chat

IS – Information Systems

LPM – Linear Probability Model

OLS – Ordinary Least Squares

OSM – OpenStreetMap

REST(-ful) – Representational State Transfer

SDK – Software Development Kit

URL/I – Uniform Resource Locator/Identifier

VGI – Volunteered Geographic Information

To my mother,

*who taught me to embrace having
more questions than answers.*

CHAPTER 1 – INTRODUCTION

An ever growing part of economic activity is facilitated by digital technology. Digital platforms, in particular, have risen to prominence and are perhaps *the* dominant approach to designing and managing digitally-enabled activity at the moment. Digital platforms are complex digital technologies that facilitate interaction between parties that would have been hard or even impossible to connect in the absence of the platform (Gawer, 2014). The interaction of large numbers of participants who are distributed in time and space enabled by platforms often generates novel outputs (Gawer, 2014; Lyytinen, Yoo, and Boland, 2016; Yoo, Henfridsson, and Lyytinen, 2010). Indeed, a wide range of economic outputs are now generated on platforms – be that in the form of new transactions, products and services, or derivative digital technology artifacts.

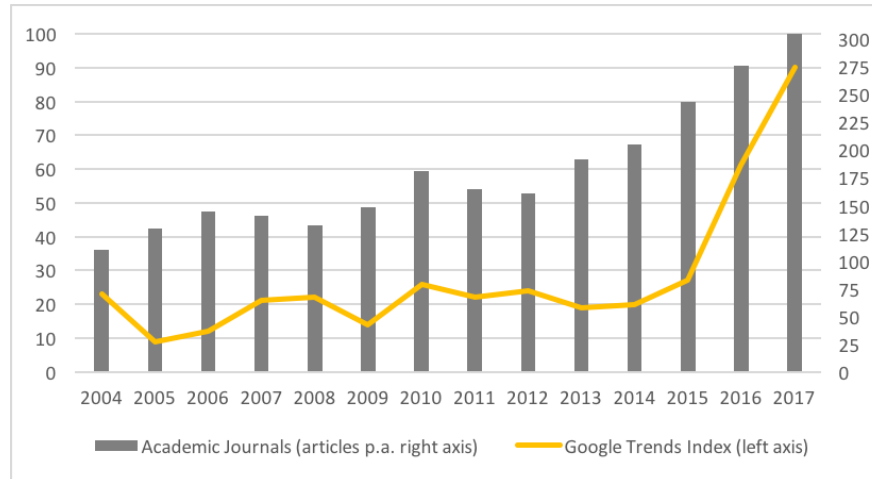
With this understanding, platforms are immensely impactful to our economic and social lives. Yet, the organising logic of connecting otherwise disconnected parties is not new. What is new, is that information technology facilitates the setup and operation of digital platforms with a substantially reduced need for physical assets (Van Alstyne, Parker, and Choudary, 2016). Many organisations have sought to leverage the opportunities offered therein and established themselves as platform businesses. For instance, in the time of writing this dissertation, seven of the ten most valuable companies globally base their operation on some kind of digital platform¹. A recent survey demonstrated the coverage of platform businesses across industries as diverse as media, software, travel and hospitality, transportation, banking, healthcare, and energy (Evans and Gawer, 2016).

Unsurprisingly, such organisations and the dynamics that characterise them have attracted a lot of attention. As a result, interest in digital platforms has proliferated, both in academe and beyond. Figure 1.1 below illustrates the number of academic publications with “digital platform” in the title or abstract (grey bars) as well as the number of worldwide searches for the term “digital platform” on the internet search

¹ Top 10 valuable companies in terms of market capitalisation: 1) Apple, 2) Alphabet, 3) Microsoft, 4) Amazon, 5) Tencent, 6) Berkshire Hathaway, 7) Alibaba Group, 8) Facebook, 9) JPMorgan Chase, 10) Johnson & Johnson; All data based on ycharts.com; retrieved 31 March 2018

engine Google (orange line). A steady increase is clearly visible over the last ten years with sharp uptake over the last four years (2014-2018).

Figure 1.1. Interest in Digital Platforms



Grey Bars: research database query; "digital platform"; 01/2004 - 12/2017; search in title and abstracts of peer reviewed articles in academic journals; included databases; BusinessSourceComplete; EconLit; JStor, ScienceDirect, Digital Access to Scholarship at Harvard (DASH)

Orange Line: Google search trends "digital platform"; 01/2004 - 12/2017; indexed popularity of global searches relative to the current high in 08/2018 (=100)

Despite this pervasiveness, our understanding of the dynamics on platforms merits expansion. How platforms facilitate novel output by enabling interaction among diverse parties is non-trivial². While platforms are often purposefully designed and managed by an operator³, platforms grow and innovate largely outside the control of any single party. The interaction between a platform and its participants often leads to non-obvious trajectories as interactions unfold in the absence of a central plan, generating novel outputs beyond the initial intent of an operator.

This introductory chapter will outline the core assumption that motivated writing this dissertation: Innovating with digital technology differs fundamentally from innovating

² indeed, the dynamics that shape activity on platforms can be paradoxical; see Lyytinen *et al.* (2018) for a detailed discussion

³ The literature offers various alternative terms, such as platform sponsor, platform designer, or platform owner. The main intention of such denotation is to demarcate the entity in control of design, governance, and architecture of a digital platform from parties using and interacting on the platform. For the remainder of this dissertation, the term "platform operator" will be used to refer to this entity. As such, the operator of a platform stands in contrast to actors engaging with the platform. This dissertation refers to such actors as "platform participants".

with physical materials and this difference is particularly observable on digital platforms. Simply put, the activities of designing, producing, and commercialising an industrial product such as a car are distinct from the activities that go into developing and maintaining a digital product such as a smartphone application. This introductory chapter provides a definition of digital innovation and relates such innovation to digital platforms. The chapter closes by highlighting the empirical setting where technological, economic, and social aspects of innovation on platforms are observable to a level of detail not easily found elsewhere.

1.1. DIGITAL INNOVATION

Our current understanding of innovation is largely rooted in theory aimed at explaining organisational activity in the industrial age. Broadly speaking, innovation is depicted as the result of actors creating, introducing, and commercialising novelty as part of market-based exchanges (Schumpeter, 1983). Central to this industrial view of innovation is the transformation of physical matter: raw materials are acquired and manipulated with the goal to yield marketable products. No wonder that significant streams of research such as industrial organisation and competitive strategy are based on the logics of transforming and exchanging physical goods among market actors (Langlois, 2007; Porter, 1980).

Three fundamental aspects of innovation characterise such an understanding (also compare Bowman, 2015; Lusch and Nambisan, 2015; Ng and Smith, 2012):

(1) *Producer – consumer divide*

The parties of the transaction are clearly demarcated. There is at least one producing and one consuming party between which an economic exchange of tradable goods is realised.

(2) *Product – process innovation distinction*

Product and process innovation are clearly separated and denote a new outcome or a new method respectively.

(3) *Innovation as deliberate organisational activity*

Innovation is viewed as an essentially plannable and controllable organisational process. As such, innovation processes are divided into

distinct steps such as identifying, selecting, and implementing⁴. Governed by basic principles of hierarchical control and division of labour within organisations, innovation is subject to managerial action aimed at the successful commercialisation of a novel product or adopting a new process.

This core logic shaped theory on innovation and its role in economic activity for most of the 20th century. Dominated by organisational perspectives on the creation, commercialisation, and implementation of information technology, the IS field has also been substantially influenced by these ideas in its view of innovation (Fichman, Santos, and Zheng, 2014; Lyytinen and Rose, 2003; Van de Ven *et al.*, 1999).

1.1.1. Definition

For the sake of this dissertation, digital innovation is defined as the “*co-creation of novel outputs through recombination of digital technology components*” (cf. Hukal and Henfridsson, 2017). As such, digital innovation is prevalent on digital platforms due to the interaction platforms facilitate. Three core aspects of digital innovation are detailed below: digital technology, recombination, and co-creation.

1.1.2. Digital Technology

Digital technology possesses properties that are unique and distinct from physical materials. These properties are at the core of the current understanding of digital innovation (Boland, Lyytinen, and Yoo, 2007; Henfridsson, Mathiassen, and Svahn, 2014; Kallinikos, Aaltonen, and Marton, 2013a; Yoo *et al.*, 2010). There are at least three properties with specific implications for organisation and innovation (Yoo *et al.*, 2010). First, digital technology involves homogenising data. Once digitised, information in digital form (bits) can be stored and transmitted irrespective of its content type by any device with computing capabilities. Second, digital technology is re-programmable. Digital bits are editable at any point in time, making digital technology malleable to changes after the fact through interaction by human and non-human actors such as other technologies. Third, digital technology is needed to create digital innovations. In other words, digital technology is characterised by self-

⁴ Models of archetypical innovation processes differ widely in the literature; we here refer to a simplified form, compare e.g., Rogers (1983)

reference, that is, it is both the result of and the basis for developing new digital technologies (see Yoo *et al.*, 2010).

The pervasiveness of digital technology and its connectivity culminates in digital infrastructures, defined as the entirety of connected, unbound, and evolving socio-technical systems that render organisational activity (Hanseth and Lyytinen, 2010; Tilson, Lyytinen, and Sørensen, 2010). These infrastructures⁵ form the foundation on which digital platforms rest (Constantinides, Henfridsson, and Parker, 2018). The unique properties of digital technology induce growth in digital technologies by fuelling two interrelated processes; digitisation and digitalisation (Tilson *et al.*, 2010). Digitisation describes the process of representing information in digital form – be that information stored in existing repositories or generating new information. This is mirrored by the process of digitalisation, i.e. the wide spread use of digital technology (Tilson *et al.*, 2010). Driven by rapid advances in developments of computing technology, the availability and affordability of performant and connective devices contribute to the ubiquity of digitally stored information. In combination, the dynamics of digitisation and digitalisation jointly enable diverse digital technology artifacts to interoperate on the basis of accessing and manipulating a common resource; digitally stored information. Digital platforms in particular draw from the abundance of digital information as well as the interoperability and connectivity of digital technology artifacts.

1.1.3. Recombination

The second important aspect of digital innovation on digital platforms is recombination. It is driven by three fundamental dynamics of digital innovation:

- (1) The separation of form and function
- (2) The separation of content and medium
- (3) Generativity

First, digital innovation benefits from the separation of form and function (Yoo *et al.*, 2010). In contrast to physical objects, digital technology allows alterations of functionality after production. Digital artifacts can be re-programmed so that the same underlying form delivers new functionality. This is possible because the semiotic

⁵ A detailed discussion is covered in Lyytinen *et al.* (2018)

functional logic is independent of the physical device executing it (Yoo *et al.*, 2010). For instance, developing, installing, and executing a new smartphone application counting the steps a user takes per day, adds new functionality to the phone without changing the underlying device. In contrast, physical objects are characterised by a tight coupling between the functions of the object and its physical composition. Physical objects by virtue of binding materials into a composite product also define a static link between forms and the functions they embody. Contrarily, accessing a subset of available resources and applying artifact agnostic changes is commonly found in digital innovation (Yoo *et al.*, 2010). This affords digital technology with the flexibility to amend its functionalities regardless of the initial design, and renders digital technology mutable depending on factors such as the application domain, user groups, or the relation with other artifacts (Henfridsson *et al.*, 2018).

Second, digital innovation is shaped by the separation of medium and content. Due to the homogenous character of digital data, content can be stored, transferred, changed, accessed, or deleted irrespective of the medium: In principle, any device with computing capability can be used to handle digital information. Equipped with networking capabilities, digital technologies enable such information handling processes across devices. This often results in the creation of vast and pervasive networks of connected digital technologies that are orchestrated to fulfil information processing tasks (Yoo *et al.*, 2012). The emerging assemblages of computing devices alter interaction with technology as, for instance, sensory technology records digital information without direct input from the user, which is then processed by a network of associated devices until eventually put to use to configure a downstream service (Yoo, 2010). Rather than confined to a predefined medium, digital content can thus be processed by a variety of devices each fulfilling different tasks.

Third, the unbundling of form from function, as well as content from medium, are amplified by the self-referential character of digital technology that can lead to generativity⁶ (Lyytinen, Sørensen, and Tilson, 2018; Tilson *et al.*, 2010; Yoo *et al.*, 2010; Zittrain, 2006). Embedded in vast networks of connected devices, newly created components of digital technology are both the result of and the basis for recombination. Heterogeneous and distributed interaction with digital technology

⁶ Lyytinen *et al.* (2018) define generativity as “the capability and related mechanisms for unbounded growth in scale and diversity of the functions and embeddedness of the infrastructure.”

artifacts thereby regularly results in non-obvious recombinant outcomes – a behaviour referred to as generativity (Yoo *et al.*, 2012; Zittrain, 2006). Generativity denotes novel unanticipated outcomes of interaction with digital technology beyond the initial design, often without deliberate planning, and in the absence of direct control through the originator of the technology (Wareham, Fox, and Cano Giner, 2014). Recent work emphasises the notion of diverse and unforeseen interactions enabled by digital technology, making generativity a crucial aspect of digital innovation.

1.1.4. Co-Creation

Finally, innovation on digital platforms is characterised by the co-creation of novel outputs. The way in which digital technology components are (re-)combined alter the way novel outputs are generated by organisations. Physical output is traditionally created through transformation of physical matter in a sequence of steps that form a firm's activity. However, the creation of output as a function of production and transaction has limitations in digital innovation. The notion of the notion of novel output as a result of co-creation by networks of actors is a cornerstone of most digital innovation studies (Lyytinen *et al.*, 2016). Rather than having been unilaterally produced, novel outputs are generated by dynamic processes of resource combination and integration (Lusch and Nambisan, 2015).

The attributes of digital technology described above lead to an increase in the density and availability of the core resource needed for digital innovation: digitally stored information. At the same time, connectivity among digital technology artifacts advances liquefaction of digital information (Lusch and Nambisan, 2015). This makes digitally stored information widely accessible and available for resource integrating activity among actors in order to configure novel digital services (Henfridsson *et al.*, 2018)⁷. Co-creation thus renders digital innovation relational. Digital innovations are created through connections among actors and artifacts rather than through isolated production. By actualising connections, novel output is created through the interaction

⁷ Note: This understanding has obvious implications for processes of value creation and capturing by organisations. However, such a discussion of “value” of digital innovation is omitted here. It should be noted that it is the view in this dissertation that the value of digital innovations does not equate to monetary value of exchange of an output. Rather, in absence of a direct monetary recompense, co-creation alludes to the integration of digital resources in a way that holds value-in-use when embedded in a digitally-enabled interaction. For detailed discussions see Henfridsson *et al.*, 2018; Lusch and Nambisan, 2015; Ng and Smith, 2012

between actors and artifacts on the basis of digital technology and its unique attributes (cf. Lusch and Nambisan, 2015).

This dynamic is particularly relevant on digital platforms where resource flows are facilitated by digital technologies. By exchanging and integrating resources through reciprocal connections, inputs for novel combinations are introduced from diverse origins across a network platform participants (Benkler, 2006; Lyytinen *et al.*, 2016). Innovation on digital platforms is thus not the result of isolated activities by one focal organisation. Instead, novel output is generated through dynamic co-creation processes of recombining technology components and interacting with diverse repositories of digitally stored information.

1.2. PLATFORMS AS MEANS OF ORGANISING DIGITAL INNOVATION

Digital innovation as characterised above is particularly prevalent on digital platforms. Platforms enable interaction among otherwise unaffiliated organisations (Jacobides, Cennamo, and Gawer, 2018). The driving force behind such interaction are connected digital technology artifacts and the activity they afford (Constantinides *et al.*, 2018; Lyytinen *et al.*, 2018). Digital platforms are layered-modular technologies whose components constitute shared artifacts that interoperate via standardised interfaces (Yoo *et al.*, 2010). As such, digital platforms facilitate the exchange of digital information that is used for diverse parties to interact on the platform and generate novel output (cf. Gawer, 2014). As a result, the unique attributes of digital technology and the logic by which the recombination of digital technology components on platforms generate novel output through co-creation, merits a reflection of the three core notions of innovation described earlier.

(1) *Producer-consumer division*

Since digital innovations are co-created by diverse actors, the distinction between producer and consumer increasingly blurs. Actors benefiting from, and using, digital innovations are often involved in their creation and development. For instance, many digital platforms generate novel output through resource exchange and integration by enabling interaction among users, external developers, and connected artifacts. Rather than the production on one end of a transactional exchange and the consumption on the other, innovation on digital platforms is better characterised by co-creation spanning design, delivery, and

commercialisation of output based on recombination of digital technology components (Lyytinen *et al.*, 2016).

(2) *Product vs. process distinction*

Digital innovation challenges the distinction between product and process innovation as both aspects are subject to constant interaction with digital technology. Digital innovation hence entangles notions of products and processes. The composition of functions and features in a digital product underlies continuous adjustment given the attributes of digital technology and the role of editable digital information as its core resource. Actualising digital innovation entails changes to the way outcomes are achieved as well as altering the outcome itself. The exchange of digital information on the basis of digital technology use describes both processes and products⁸ (Lyytinen *et al.*, 2016; Nambisan *et al.*, 2017).

(3) *Innovation as deliberate organisational activity*

Distinct stages as in a controlled, linear organisational process are not clearly discernible in digital innovation. Instead of a discrete sequence of steps, the mechanisms that generate digital innovations are more akin to continuous, iterative cycles. Such iterations are serendipitous to a degree that makes deliberate process phases unrecognisable (cf. Van de Ven *et al.*, 1999). Rather than a stringent sequence of steps yielding an end result, sufficient iteration eventually culminates in stable coherent set of functions that form novel output at a given point in time. This stabilisation, however, is only the basis for the next iteration and refinement of the output, inducing a new version and potentially extending functionalities by including new actors.

The above descriptions notwithstanding, our understanding of exactly how platforms generate innovations by enabling interaction among diverse parties is still developing. The distributed and combinatorial nature of digital innovation outlined above marks a challenge for information systems researchers. Digital innovation requires an understanding of organisational activity with an emphasis on constant interaction across boundaries (Yoo *et al.*, 2012). On digital platforms, innovations are generated

⁸ This understanding is not unique to digital innovation. In fact, management and organisation scholars have long disputed the distinction between innovation as either a product or a procedure e.g., compare Van de Ven *et al.* (1999)

by enabling activity that routinely transcends teams, firms, industries, and markets. The growth of digital platforms is characterised by incorporating such novel outputs in the architecture and operation of digital platforms. In order to understand these aspects of digital platforms, attention is needed towards the supra-organisational dynamics emerging from the interaction on platforms (Lyytinen *et al.*, 2016). This dissertation aims at doing exactly that. The overarching line of inquiry can be summarised by the question; “*How do digital platforms grow and innovate?*”.

1.3. EMPIRICAL SETTING

The dissertation addresses the above question in the context of the open geo-data platform OpenStreetMap (OSM). OSM, the platform, is an open source software artifact which produces the assets of OpenStreetMap, the geo-spatial data project, available both via programmable interfaces and via the browser (OpenStreetMap.org). In operation since 2004, ‘the Wikipedia of Cartography’ is considered the world’s largest community-driven mapping project on the web with over 4.5 million registered users who to date contributed and edited more than 4.2bn geo-spatial data points. OpenStreetMap is not only free of charge, it is also “free of restrictions that hinder the productive use of the data” (Ramm, Topf, and Chilton, 2010; p. 3).

OSM is a well-suited research setting for this dissertation as it provides the opportunity to study various aspects of innovation on digital platforms. In this setting, technological, economic, and social aspects of digital innovation are observable to a level of detail not readily found elsewhere.

The platform consists of a simple set of functionalities aimed at providing vast amount of geo-spatial data in the form of geo-locations and meta-data descriptions. The data is “volunteered geographic information (VGI)”, which is typically contributed by individual users (Goodchild, 2007). Users upload geo-spatial data points, for instance, in the form of GPS traces or by interpreting aerial imagery using various editing software tools⁹ (Ramm, 2015; Ramm *et al.*, 2010). This GPS data is the foundation of every object in the OSM database (Mooney and Corcoran, 2012). Data objects are simple point geometries that can be grouped together to form lines or polygons in order to represent all sorts of geographical features – natural or man-made. Stored geo-data can then be annotated with semantic information through free-text labels – so

⁹ For an overview see https://wiki.openstreetmap.org/wiki/Comparison_of_editors; accessed 10-02-16

called tags. OSM tags complement the data model¹⁰ and take the form of “*key = value*” to describe real-world objects in great detail (Ramm, 2015). On the level of content, the platform facilitates interaction between a database and participants contributing, editing, and retrieving geo-spatial data.

The case of OSM also offers the opportunity to study technology architecture and design of digital platforms. Software used to operate OSM includes, among others, a frontend web application, several web application programming interfaces (APIs), data editing tools, as well as highly specialised tools to handle geo-spatial data (Hukal, 2017; Hukal and Eck, 2016; Ramm *et al.*, 2010). These tools are self-contained software programs that extend the database functionality of the core platform (geo-spatial database). External application developers wanting to use geo-spatial data handling capabilities in their own services thus have several software tools available in a mix-and-match manner. In this view, the OSM platform can be conceptualised as an extensible code base whose functionality is augmented by modular software components (e.g., Baldwin and Woodard, 2009). Over time, the platform’s technical architecture has been adapted and extended to include many technical solutions provided by external developers such as leaflet.js (a library for interactive maps), Rails (a web development framework), or Mapnik (a rendering engine). In this sense, the platform is subject to the organisation and management of complex and dynamic ecosystem of technologies (Jacobides *et al.*, 2018; de Reuver, Sørensen, and Basole, 2018). Therefore, researching OSM allows the investigation of technical aspects of platforms such as architecture, functionality, or design of the platform and its add-on software modules.

Furthermore, advantageous access to technical development data of OSM invites investigations into technology architecture and design of platforms. The source code of OSM software tools used to interact and extend the core database functionality is hosted on GitHub to coordinate development work on the codebase. GitHub offers a web-based version control system on the foundation of the distributed version control software Git¹¹ and lets users share, propose, and discuss software code (Dabbish *et al.*, 2012). In the case of an open source software project such as OSM, any GitHub user

¹⁰ The term data model in the context of OSM usually refers to the logic of storing GPS data as points, lines, or polygons which OSM contributors refer to as nodes, ways, or relations.

¹¹ For details see e.g., Chacon and Straub (2018), Pro Git – Everything you need to know about Git, 2nd Edition, APress, available under CC-SA-30 at <https://git-scm.com/book/en/v2>

may view and access its codebase and propose source code changes (Tsay, Dabbish, and Herbsleb, 2014). GitHub makes the content of source code changes together with a number of metadata metrics publicly available¹².

Lastly, the case of OSM is of commercial interest. The data held by OSM as well as the various capabilities to handle geo-spatial information are immensely popular among external application developers and service providers. Hundreds of commercial and non-commercial web services draw from OSM data and related data handling capabilities in their configurations¹³. For instance, Craigslist, Wikipedia, Garmin, Citymapper, or Foursquare all use geo-data provided by OSM as part of their products (see Ramm *et al.*, 2010 for details). As such, OpenStreetMap stands in direct competition with popular proprietary products such as GoogleMaps, TomTom, or Nokia HERE. Competition among these geo-spatial data services centres around offering up-to-date geo-data, at the highest accuracy, across a wide range of devices and operating systems (Parsons, 2013).

In summary, the OpenStreetMap platform is superbly positioned for research seeking to explore technical, economic, and social aspects of innovation on digital platforms.

¹² Chapters 5 and 6 make extensive use of development data collected from GitHub and will provide more detail on structure, access, and use of such information.

¹³ See https://wiki.openstreetmap.org/wiki/List_of_OSM-based_services

1.4. STRUCTURE OF THE DISSERTATION

The main part of the dissertation is made up of three empirical studies. Each study investigates one specific notion of innovation on digital platforms by focusing specific dynamics on the OpenStreetMap platform. Chapter 4 addresses the question of how platform operators can guide participant interaction on platforms in order to stimulate generativity of the platform. Chapter 5 reflects on design principles of platforms in light of the unique aspects of digital technology. Chapter 6 reports on participant coordination on digital platforms when interface components can no longer provide stable interaction with the platform.

The dissertation proceeds as follows. The second chapter reviews relevant studies on digital platforms in the management information systems literature. In so doing, the section focuses on major lines of inquiry and highlights potential contributions to the study of digital platforms. Subsequently, chapter three outlines the analytical approach that ties together the work undertaken in the empirical part of this dissertation. Chapters four through six present the empirical studies on growth and innovation phenomena on digital platforms. Chapter seven summarises the findings and points to future research.

CHAPTER 2 – LITERATURE REVIEW

This section reviews information systems literature on digital platforms. The section serves two main purposes. First, this chapter situates the dissertation in the ongoing debates about digital platforms by providing an overview of current themes and topics in the wider field of information systems. Second, the chapter introduces concepts and arguments that motivate the work undertaken in this dissertation in preparation of the subsequent empirical studies.

To avoid repetition, the literature review in this chapter is deliberately general. The empirical studies presented in this dissertation each draw from their own literature review in order to make specific and separate contributions to the field. As such, the literature reviewed in the empirical studies emphasises the focus of the respective study. The empirical studies deal with core concepts in more detail in order to concentrate on the respective contribution target. Whereas the concept of platform generativity is the subject of the first empirical chapter, platform modularity and platform design will be at the centre of the second study. Lastly, the third chapter will deal with boundary resources and coordination among platform participants. It is therefore important to note that the present chapter highlights topics that are important to research on digital platforms in general, yet not all topics covered in this review are necessarily of chief concern for the dissertation.

The following section frames the literature on digital platforms by delineating major streams of research and highlighting salient studies within them. The chapter will close by outlining lines of inquiry and potential contributions to the literature on digital platforms.

2.1. MAJOR STREAMS IN RESEARCH ON DIGITAL PLATFORMS

Essentially, the literature on digital platforms can be divided into four major streams of research. While interrelated and partially overlapping, the studies in each of these streams have distinctly recognisable foci in their view on platforms as well as their main interest of investigation. Table 2.1 below gives an overview over the major streams of research.

Table 2.1. Major Streams of Research on Digital Platforms		
Stream of Literature	View on Platforms	Salient Studies in IS Literature
Platforms as Marketplaces	<p>Platforms are multi-sided marketplaces enabling economic transactions to take place among otherwise unconnected entities.</p> <p>Platforms grow and innovate by stimulating and maintaining network effects between demand and supply side entities.</p>	<p>Boudreau (2010; 2012) Eisenmann <i>et al.</i> (2011) Niculescu <i>et al.</i> (2018) Parker <i>et al.</i> (2017) Rietveld and Eggers (2018)</p>
Platforms as Modular Systems	<p>Platforms are modular technologies comprising a stable core of executable software code that is extended by add-on functionality – often developed by third parties.</p> <p>Platforms grow and innovate by managing the technical design of platform core, boundary, and periphery components in order to stimulate unanticipated growth trajectories.</p>	<p>Baldwin and Clark (2006) Baldwin and Woodard (2009) Eaton <i>et al.</i> (2015) Ghazawneh and Henfridsson (2013) Kazan <i>et al.</i> (2018) Karhu <i>et al.</i> (2018) Tiwana <i>et al.</i> (2010) Woodard <i>et al.</i> (2013) Yoo <i>et al.</i> (2010; 2012)</p>
Platforms as Ecosystems	<p>Platforms are ecosystems of shared technologies.</p> <p>Platforms grow and innovate through three forces akin to biological evolution; heredity, mutation, and adaptation.</p>	<p>Um <i>et al.</i> (2013; 2015) Um and Yoo (2016) Yoo (2012) Zhang <i>et al.</i> (2015)</p>
Platforms as Socio-Technical Systems	<p>Platforms are socio-technical assemblages of values, norms, and practices that render social activity.</p> <p>Platforms grow and innovate by reinforcing social interaction through increased use of the platform.</p>	<p>Alaimo and Kallinikos (2017) Clark <i>et al.</i> (2014) Helmond (2015) Plantin <i>et al.</i> (2018)</p>

2.1.1. Platforms as Marketplaces

First, a popular stream in research on digital platforms is informed by the tradition of *economics of information systems*. Here, views on platform growth take as their point of departure the notion of platforms as multi-sided marketplaces facilitating economic interactions between actors that would have been hard or impossible to connect in absence of the platform (Gawer, 2014; Rochet and Tirole, 2003). While economics of

IS has in the past studied a range of different kinds of platforms (such as firm, supply chain, or product platforms), the notion of an industry platform that facilitates interaction across organisational boundaries is the most dominant notion represented in information systems studies (compare Gawer, 2014). As such, platforms grow as long as the enabled economic exchanges add value to the various actors on the platform (Boudreau, 2012; Ceccagnoli, Forman, and Huang, 2012).

In this view the notion of network effects aimed at achieving critical mass in adoption is fundamental (Eisenmann, Parker, and Van Alstyne, 2011; Evans, 2009). Network effects reflect complementarities in demand and supply such that attracting actors from both sides is crucial for platform growth and participation of either side is dependent on the other (Gawer, 2014; Parker, Van Alstyne, and Jiang, 2017; Rochet and Tirole, 2003). Managing platform growth is therefore seen as balancing complementary demand and supply by enabling interaction among both the two sides (e.g., Gawer and Henderson, 2007). Studies in the economics of IS tradition are thus interested in managing the network of actors and investigating different pathways to platform growth and innovation.

In order to remain competitive, platforms strive to increase economic surplus for actors involved on the platform. In this view it is imperative to stimulate complementarities in demand and supply in a way that results in growth and innovation (Evans, 2009; Gawer and Henderson, 2007). New entrants are thereby often confronted with what is colloquially referred to as the ‘chicken and egg problem’ (e.g., Boudreau, 2012; Gawer, 2014). That is, platform operators face the challenge of simultaneously attracting parties from both the demand and the supply side to the platform. In the extreme case of a new entrant, the problem is exacerbated by a seemingly non-existing benefit for participants of committing to a platform with no party to interact with. Evans (2009), for instance, conceptualises the strategies available to new entrants seeking to grow a platform by means of identifying what he calls ‘inflection points’ aimed at creating critical mass (Evans, 2009). In the same sense, Rietveld and Eggers (2018) argue that this also presents a driver for growth for platform operators as platform participants on the supply side benefit from engaging with platforms that are home to early adopters on the demand side. They find that suppliers of platform complements enjoy higher sales on platforms predominantly used by early adopters and that the share of sales changes as an increasing

heterogeneity in the platform's demand structure shifts to late adopters (Rietveld and Eggers, 2018).

The notion of managing multiple complementarities on the platform is firmly rooted in the economic view, and most derivative platform strategies draw from its logic. In essence, the view of platforms as marketplaces informs three mechanisms of platform growth in the literature.

First, managing openness of the platform. Here, several studies explicate pathways to platform growth by means of regulating the degree of market access granted to actors. Platform openness has far reaching implications for platform growth and innovation. As highlighted by Boudreau (2010), the effects to derivative innovation on platforms are contingent on the extent to which platform operators decide to grant access to platform markets and their resources (Boudreau, 2010). In a subsequent study, Boudreau points out that platform openness determines the diversity of product availability on a platform. He argues that product availability is a function of the number and the degree of specialization of suppliers engaged on a platform (Boudreau, 2012). Similarly, Niculescu *et al.* (2018), explore the conditions under which opening up platform resources for use by complementary and competing actors afford benefits to the platform operator as adoption of the platform increases (Niculescu, Wu, and Xu, 2018).

Openness can have advantages for all actors on the platform as openness and access affect the capabilities shared and gained for all platform participants. Ceccagnoli *et al.* (2012), for instance, demonstrate how interfirm collaboration on platforms benefits peripheral actors and platform operator alike if the capabilities gained through platform engagement complement peripheral actors' activities (Ceccagnoli *et al.*, 2012).

The second mechanism involves standardisation. Openness and standardisation are related yet distinct concepts. Opening a platform does not necessitate the use of a standard. Conversely, participating in a standard does not automatically translate into openness of platforms. Varian *et al.* (2004) suggest establishing, negotiating, or participating in standards as fundamental levers for stimulating network effects and thus managing platforms (Varian, Farrel, and Shapiro, 2004). The common view is that standards help the networks of actors on platforms grow by means of reducing the

effort needed for them to participate and join a platform (Brynjolfsson and Kemerer, 1996; McIntyre and Srinivasan, 2017) Standards thereby support the platform by what others have referred to as a signal (Ho and Rai, 2017). In their study of continued engagement of complement contributors on a platform, Ho and Rai (2017) find that standardised accreditation of contributions, for instance, bolsters the rate with which participants engage on the platform.

The third mechanism involves pricing. In the context of digital platforms, pricing can either refer to a tactic aimed at inducing platform growth by attracting users or facilitating the consumption of complementary products and services offered on a platform. Pricing is related to versioning of information goods, using the platform to market complementary products and services according to individuals' willingness to pay (Economides and Katsamakas, 2006; Parker and Van Alstyne, 2005; Shapiro and Varian, 1998). Parker and Van Alstyne (2005) argue that the interaction between different sides of platform markets can induce network effects, thereby helping the platform grow. In line with others (e.g., Evans, 2009), they demonstrate that pricing in the form of discounting access for one party is a potent lever for platform operators seeking to grow their platform (Parker and Van Alstyne, 2005). How platform operators set prices for platform access over time has been formalized by Parker et al. 2017. They demonstrate that platforms fuel growth and innovation by managing how and when platform participants engage with the platform (Parker *et al.*, 2017). In line with this, it has been suggested that the combination of pricing and offering derivative versions of complementary products is a viable competitive platform strategy. In what has been called "enveloping", platform operators can bundle offerings on their platform such that utility of rival platform use is subsumed (Eisenmann *et al.*, 2011).

2.1.2. Platforms as Modular Systems

A second major stream in digital platform research is influenced by *technology and IS tradition* and regards platforms as modular technology systems. Research in this tradition seeks to explain the conditions under which digital platforms grow and innovate, conditional on their technological design. This stream has studied a range of complex information technologies such as the Internet (Zittrain, 2006), and digital infrastructures (Henfridsson and Bygstad, 2013; Tilson *et al.*, 2010) and has had significant influence on research on platforms (Gawer, 2014; Tiwana, Konsynski, and Bush, 2010)

Studies in this stream tend to be informed by modularity of technology (Simon 1962, Schilling 2000). As such, platforms are thought of as modular systems enabling complex interaction of relatively independent sub-system components by means of shared artifacts and standardized interfaces (Yoo *et al.*, 2010).

The conceptualization of digital platforms as consisting of a stable software codebase whose functionality is extended by external add-on modules, often developed by third-party developers is core to technology and information systems is (de Reuver *et al.*, 2018; Tiwana *et al.*, 2010). Such platforms provide a technology foundation via a set of relatively stable functionalities with low variety that act as the base of the digital platform (Yoo *et al.*, 2010). This platform base provides a set of functionalities and resources that are available for interaction with external actors.

Facilitating interaction between external actors and platform core modules is a key mechanism for growth and innovation on digital platforms (Baldwin and Woodard, 2009). The desired objective for platform operators is thereby to have external participants add functionality and thus extend the core of the platform through component reuse.

A cornerstone in this stream of research is the view that platform architecture is the key lever available to platform operators in facilitating platform growth and innovation (Eaton *et al.*, 2015; Hanseth and Lyytinen, 2010; Henfridsson *et al.*, 2014; Tilson *et al.*, 2010; Wareham *et al.*, 2014). As such, platforms grow and innovate if the technical architecture facilitates interaction between platform core and periphery in a way that recombination of platform resources leads to the creation of novel products and services (Lusch and Nambisan, 2015; Yoo *et al.*, 2010). Innovation on

platforms is typically seen as a self-reinforcing process of using, appropriating, and re-combining resources made available by the platform operator and affiliated modules. For instance, in their paradigmatic typological analysis, Ghazawneh and Henfridsson (2015) classify platforms according to their ability to produce numerous and general applications if resources are openly and flexibly available to external parties (Ghazawneh and Henfridsson, 2015). They suggest that design decisions along a spectrum of control and openness have implications for the scale and scope of innovative activity that a platform is expected to generate (Ghazawneh and Henfridsson, 2015).

A common notion in studies in this stream is the attention given to unique attributes of digital technology and the consequences for innovation on digital platforms. As a consequence, a large part of the discourse in this stream addresses the challenge of stimulating platform growth and innovation while also exercising control over the platform (Hanseth and Lyytinen, 2010; Svahn, Mathiassen, and Lindgren, 2017; Tilson *et al.*, 2010). Wareham *et al.*, (2014), for example, explicate the intricate dynamics arising from platform interactions and indicate architectural designs as a governing mechanism to resolve tensions between the competing interests of actors on platforms. They point out that platform operators need to balance desirable and undesirable variation when attempting to grow the platform (Wareham *et al.*, 2014). One course of action suggested by Wareham *et al.* (2014) is the implementation of governance control into the participative architecture of the platform. For example, they argue that selective modularisation of platform components or certification of external contributions are mechanisms by which the platform operator can increase control over third party activity on the platform (Wareham *et al.*, 2014).

Other scholars have also focused on structural arrangements in order to manage tensions between the platform operator, who wants to retain control, and heterogeneous third-party developers, who want to pursue their own ideas. Studies have explored the ways in which the platform operator governs the interaction between platform core and periphery.

For instance, attention was given to the design of components on the platform boundary. So called “boundary resources” have been highlighted as important strategic components capable of steering what interactions are permissible on a platform and how by enabling connection and communication across the platform

boundary (Ghazawneh and Henfridsson, 2013). Work on boundary resources has highlighted how acts of ‘resourcing and securing’ the platform contribute to its growth (Ghazawneh and Henfridsson, 2013). The key task of boundary resources is therefore the continuous-iterative regulation of what resources are available for interaction between core and periphery and how. Expanding platforms by offering a new resource accompanied by the simultaneous establishing of rules for interaction manages platform growth by means of facilitating interaction with the platform core (Eaton *et al.*, 2015; Ghazawneh and Henfridsson, 2013). For instance, Eaton *et al.* (2015), demonstrate how the Apple iOS platform evolves as modules at the platform boundary are shaped by heterogeneous and distributed actors generating results largely outside of the control of the platform operator (Eaton *et al.*, 2015). In reaction to tussles arising from initial unilateral designs implemented by the platform operator, they observe accommodation and resistance of diverse actors’ interests changing boundary resources over time (Eaton *et al.*, 2015). This process contributed to the expansion of the platform albeit in contradiction to the initial intend of the platform operator.

Governing activity within and across platforms through boundary resources thereby also offers a powerful lever for platform strategy. In a recent study, Karhu *et al.* (2018), demonstrate how the design of resources on the platform boundary can be decisive for how platform operators are able to compete. In their study of “platform forking”, they identify that exploiting as well as defending digital platforms is controlled via the openness of boundary resources such as interfaces, licenses, or framework agreements (Karhu, Gustafsson, and Lyytinen, 2018).

Others have focused on the interaction between technological designs and performance of platforms. As such, attention has been given to the ability of platforms to evolve contingent on their architectural designs. For instance, under a clear influence from modular systems theory, Tiwana argues that platform attributes such as malleability and plasticity (i.e. the ability of a platform to incorporate change) are driven by the architectural design decisions made in set up and operation of the platform (Agarwal and Tiwana, 2015; Tiwana *et al.*, 2010). He argues that application developers architecturally leverage the capabilities offered by the platform operator. It is through this “fine-tuning” that applications align their architectural set-up with a focal platform and realize performance benefits in the competition with other applications (Tiwana, forthcoming). Others found similar dynamics and theorise that the architectural set-up of a platform has consequences for participant engagement

with the platform. As Cennamo *et al.* (2018) demonstrate that decisions by complementors to engage with a platform or not are influenced by the effort required to align with a platform's architectural complexity (Cennamo, Ozalp, and Kretschmer, 2018).

With a view on the platform operator, Kazan *et al.* (2018) argue that the architectural design of digital platforms corresponds to the underlying business logic of the party that runs the platform. In a comparison of multiple payment platforms, they theorise that platforms compete by configuring the interaction of architecture and economic activity of the platform (Kazan *et al.*, 2018).

Yet another angle investigates design decisions in relation to enabling or constraining platform performance. One key challenge for platform operators is to attain the ability to rapidly respond to unanticipated changes. In that vein, Woodard *et al.* (2013), highlight how inertia in a platform's technical architecture interacts with their competitive performance. What they term "design capital" translates into the flexibility of a platform that an operator can utilise conditional on past design decisions as reflected in technical debt or available technological options (Woodard *et al.*, 2013). Others explore the interplay between technology options and technology debt further and turn to the challenges of managing digital platforms within organisations (Rolland, Mathiassen, and Rai, 2018). Rolland and colleagues find that the architecture of digital platforms generally increases the complexity of how technical options and technical debt interact. Challenges like this have implications for the way in which organisations approach the adoptions of digital platforms and how they interact with newly affiliated actors. Svahn *et al.* (2017), for instance, find concurrently existing concerns among managers struggling with the decision to incorporate digital platforms in firm activities such as product development (Svahn *et al.*, 2017).

2.1.3. Platforms as Ecosystems

The third stream of digital platform research is a recent development informed by complexity theory and influenced by the vocabulary of evolutionary biology. Platforms in this view are understood as ecologies of technologies bound together by shared technology use (Um and Yoo, 2016). Studies in this stream view platform growth as a process that builds up a pool of shared elements, while platform innovation is the subsequent access and recombination of these elements to create novel platform possibilities (Um *et al.*, 2013; Yoo *et al.*, 2010).

The information systems field has developed a great interest in complexity theory (e.g., Benbya and McKelvey, 2006; Merali, 2006; Tanriverdi, Rai, and Venkatraman, 2010). The non-trivial interaction between numerous parts of a complex system are said to give rise to emergent phenomena (Frenken, 2006). The underlying idea is often to treat innovation as an emergent property. Equally, those studying digital platforms have drawn on a complex adaptive systems view. In the language of complex adaptive systems, emergence is the macro level outcome of micro-level interactions (Anderson, 1999). The information systems field has embraced that idea in studies on digital platforms in attempts to explain how innovations on platforms (such as novel applications) come about as the result of interaction between largely independent parts of a complex technical system such as platforms.

While some overlap with modularity (compare Simon, 2002) exists, IS studies in this stream are distinct. For instance, the term ‘ecosystem’ is widely used in the literature on technology innovation management and strategy (e.g., Adner and Kapoor, 2016; Jacobides *et al.*, 2018). Here, modularity and the understanding of complexity therein is a clear influence, and their terminology sometimes overlaps with studies on information systems. However, it is important to note that IS studies in this stream are distinct as they seek to understanding platforms through the actual make-up of platform features. Often, scholars in this stream look at platform components as genes that are catalysts for mutation and innovation (Um *et al.*, 2013). Informed by complex adaptive systems thinking about technology evolution (Fleming and Sorenson, 2001), such views recognise the influence of adaptation to environmental dynamics on how technological components are appropriated. Key to such evolutionary views is the incorporation of environmental feedback to attain a locally optimal adaptation – a

process that is seen as a core driver of change in technological structures (Frenken, 2006; Kauffman and Macready, 1995). This highlights the notion that interaction with a system's environment influences how technological components are appropriated over time (Kauffman and Macready, 1995). Murman and Frenken (2006), for instance, use the example of software platforms when they argue that the emergence of dominant designs is shaped by technological, industrial, and economic dynamics in the environment of a technology (Murmman and Frenken, 2006).

In the same vein, information systems researchers tie in evolutionary dynamics to describe platform growth and innovation. The utilised underlying framework is often that of evolutionary forces of heredity, mutation, and adaptation in explaining innovation on platforms and the ecosystem of actors they attract. As such, radically different combinations occur as part of an evolutionary trajectory of a platform if technology parts are accessible for recombination (Um *et al.*, 2013; Zhang *et al.*, 2015). With interacting attributes as their point of departure, platforms grow and innovate by affording the ability to evolve through means of passing on successful combinations to subsequent generations. Lastly, mutation reflects the idea that variation of existing parts ensures a well-adapted fit with the environment. This conceptualisation of platform evolution has been adopted in IS research (Um *et al.*, 2013; Zhang *et al.*, 2014) by tracing “genotype” attributes across populations of related technologies. In their study of web mashups, Zhang *et al.* (2015), for example, explore the recombination of existing modular components across different types of evolutionary paths of a platform (Zhang *et al.*, 2015). Their phylogenetic analysis suggests that the trajectories of platforms differ depending on whether they provide specialised niche services or address more diverse use cases. Similarly, Um *et al.* (2013) show how APIs are “basins of attractors” in a platform and that this foundation is akin to “genes” of a platform. They argue that this is an important source of growth and innovation on platforms as it broadens the design space by making resources available for wide-scale recombination. In a subsequent study, Um and Yoo (2016) demonstrate how digital platform ecosystems continue to evolve and follow novel trajectories upon the addition of external components. In their detailed, longitudinal account, they investigate how new external components, over time, become part of the technology core of a platform. This in turn, they argue, is the foundation upon which a next interaction of technological development is based, thereby giving rise to ever more innovation (Um and Yoo, 2016).

2.1.4. Platforms as Socio-Technical Systems

Yet another stream of research on digital platforms aligns with the socio-technical tradition in information systems research. Here, scholars conceive of platforms as hybrid social and technical infrastructures that simultaneously enable and constrain social activity as they “variously shape user platform involvement and participation” (Alaimo and Kallinikos, 2017). In this view, the focus of investigations is the role of platforms as mediators of social activity. The centre of attention thereby often rests on the control platforms exert by means of designing and deploying mechanisms that render social activity the integral part of platform operations. Platforms are not merely the stage on which social activity unfolds, but rather, platforms are socially and technically complex entities representing the values, norms, and practices of the actors involved in their design and ongoing co-creation: developers, users, regulators, investors and so forth (cf. Helmond 2015). Special attention is thereby given to the use of modern information technology and the influence that has on social interaction.

A focal point in this research stream is the emergence of platform intermediaries and the consequences for the social activity conducted on these platforms. Focussing on the role of platforms in shaping technology-mediated everyday social interaction, a large share of research concentrates on the context of social media: “Wikipedia, YouTube, and Facebook provide notable examples that show how individual or collective pursuits are tied to platforms” (Kallinikos *et al.*, 2013; p.367). For example, Helmond (2015) investigates how social interaction on the web is mediated by the advertising business model of social networking platforms that thrive on the accumulation of personal data to fuel their commercial operations. Helmond argues that the modularisation of digital technology artifacts is a key driver for what she calls “platformization of the web” (Helmond, 2015). Only through the flexible interaction of platform core and periphery modules are advertising-based platform business models viable at all, so Helmond. Users relying on the interfaces provided by actors such as large social media platforms are traceable beyond the platform through tokenisation. This in turn affects which peripheral suppliers on the platform interact with said users and how, once they return to the focal platform. In a similar vein, Alaimo and Kallinikos (2017) frame the processes involved in computing everyday interaction on such platforms. They investigate how patterns of individual and

collective use on platforms is rendered into objects of interest to the provider of social media platforms and associated business stakeholders (Alaimo and Kallinikos, 2017). They reconstruct the processes platform operators follow to compute and aggregate metrics (such as shopping behaviour and taste profiles) from social interaction with and on platform technology. Feedback from these processes, they argue, is what fuels platform growth through the continuous development of additional features on platforms that follow the same goal; collect, analyse, and utilize ever more user data via the platform (Alaimo and Kallinikos, 2017).

Clark *et al.* 2014 address the question of socio-technical forms of agency that arise through the infrastructural mediation of information creation and dissemination on platforms (Clark *et al.*, 2014). They argue that functions, resources, and connections available on digital media platforms constitute infrastructural environments in which user activity is embedded. These infrastructures in turn facilitate communicative actions such as storytelling, creating, and exchanging of narratives. This, the authors find, is ultimately bestowing agency to users through their interaction with one another on digital platforms (Clark *et al.*, 2014).

Building on the notion of hybrid agency, Plantin *et al.* (2018) demonstrate how the underlying logic of technology modularity contributes to the creation of a community of practice through platform usage. In their study of an academic data sharing platform, they suggest that splintered fractions of the community of scholars converge as an upcoming platform imposes a novel set of practices, values, and norms (Plantin, Lagoze, and Edwards, 2018). Complying with the intended actions on the platform inadvertently increased homogeneity across the community of scholars as all parts of the group increasingly subscribed to the activity expected by the platform, thus binding clustered camps in the community together via the core logic of the platform.

2.2. POTENTIAL CONTRIBUTIONS

This section derives areas for potential contributions from the literature on digital platforms. In particular, three aspects that illustrate the discourse across the dominant streams of literature on digital platforms are highlighted; (i), the detailed empirical study of digital platforms, (ii) the investigation of emergent phenomena such as generativity, (iii) the reflection on the intellectual traditions informing technology design and organisation. Past scholarship has advanced our understanding of digital platforms and the dynamics inherent therein. In so doing, these studies also motivate new questions and open the field for further contributions to understanding the growth and innovation of digital platforms.

2.2.1. Detailed Empirical Accounts

Paying attention to the design, architecture, function, and use of technical artifacts is core to the field of information systems. Careful reading of the literature reveals a need for more detailed empirical accounts of digital platforms. It is apparent from the literature that the amount of highly detailed empirical work on digital platform phenomena is limited to a handful of studies. As a consequence, de Reuver *et al.* (2017) call for “data-driven approaches and research designs” to advance scholarship on digital platforms. They claim that the explosion of tools and methods available for the collection, analysis, visualisation, and reporting of data presents a promising opportunity for researchers on digital platforms. While a push towards rich empirical accounts of digital platform phenomena – including innovation, architecture, and governance – is discernible, a diffusion into the mainstream of top-level research has only just started. Recently, Constantinides *et al.* (2018) outlined pressing issues for further research on platforms across themes such as architecture, governance, and competition (Constantinides *et al.*, 2018). Many of these can be advanced by attending to platforms in thoughtfully designed empirical studies that leverage the level of detail that digitally stored information can provide. For instance, they highlight the need for “a new mirroring hypothesis” as well as research into “growth and scaling of platforms” – both of which are addressed explicitly in this dissertation.

2.2.2. Generativity

A pervasive debate in research on digital platforms centres around the concept of generativity. Generativity generally describes a platform's ability to generate new and unanticipated outputs without direct input from the platform owner (cf. Wareham *et al.* 2014; Zittrain 2006). Much of the research of generativity is informed by the idea of digital technology enabling interaction between heterogeneous and distributed actors whose interaction among each other and with digital technology artifacts brings forth unanticipated changes in information systems (Henfridsson and Bygstad, 2013; Lyytinen *et al.*, 2018; Tilson *et al.*, 2010). A popular example of this kind of behaviour is the evolution of the Internet (Yoo, 2015; Zittrain, 2006). The various technologies comprising the Internet's underlying infrastructure are receptive to complex interactions among people and other technologies (Kallinikos, Aaltonen, and Marton, 2013b). As a result, the products and services generated with and on the internet have expanded manifold in a manner of an "exponentially expanding ecosystem" (de Reuver *et al.*, 2018).

Generativity is especially pervasive in the technology and IS stream, but the notion is represented across studies on digital platforms. While the idea does not originate in the platform space, it has certainly sparked a lot of attention among scholars interested in platforms. For instance, in the economics of IS literature, the triggering of network effects between two complementary sides of a platform is seen as the source of a high number and a high diversity of actors and products on a platform (Boudreau, 2012). The complexity and IS stream explicitly focuses on the mechanisms and attributes of generative innovation in complex platform ecosystems (e.g., Um and Yoo, 2016). In the socio-technical stream, generativity is less a subject of research and more an explanation of how and why social behaviour constantly evolves when mediated on digital platforms (Kallinikos *et al.*, 2013a).

The interest in generativity remains most salient in the technology and IS stream. Indeed, much of the vocabulary used across the digital platform space has originated in this stream. Across all streams of research, the debate on generativity is characterised by the lack of detailed empirical accounts. While some scholars strive to theorise generative dynamics from sophisticated empirical designs (Um, Eck, Zhang), very few top journal articles present empirical studies dedicated to platform generativity. Some articles do exist, yet they are few (e.g., Eaton, 2012; Svahn *et al.*,

2017), and most remain conceptual in nature (Kallinikos *et al.*, 2013b; Lyytinen *et al.*, 2018), do not theorise about digital platforms specifically (Henfridsson and Bygstad, 2013), or address generativity tangentially (Parker *et al.*, 2017). Here, this dissertation will aim at extending the platform literature by attending to generativity of digital platforms explicitly.

2.2.3. Design and Organisation of Digital Technology

It is clear that much of what has been written on digital platforms is informed by modularity. Indeed, the modular systems paradigm permeates the studies presented in the preceding section. For empirical studies on digital platforms, this is an opportunity and a challenge at the same time. It is an opportunity as modularity provides a framework with which digital platform phenomena can be investigated while remaining tethered to an intellectual framework that spans decades of research across the social and technical sciences (Campagnolo and Camuffo, 2010; Schilling, 2000). This opens avenues for researchers to empirically investigate modular structures and the organising principles therein as the make-up of digital platforms. On the platform-level of analysis, this invites studies on the design of platform components facilitating interaction as a driver of platform growth and innovation. On a supra-platform-level of analysis, modularity can inform investigations about the interaction between complex modular systems with regards to the activity and dynamics rendered on digital platforms (see e.g., Jacobides, Cennamo, and Gawer, 2018)

But equally, modularity's prevalence presents a challenge to digital platform researchers. It stands to reason that platforms are not merely like any other modular system. As others have argued, digital platforms differ from most modular systems studied in the past due to the unique attributes of digital technology (Yoo *et al.* 2010). Applying template-like expectations and adopting modularity's assumptions therefore risks misrepresenting what makes digital technology distinct from physical material. Here the empirical study of growth and innovation on digital platforms has room for meaningful contributions. Digital technology marks a departure from the nested and fixed product architectures prevalent in most physical modular systems. Instead, digital technology is characterized by layered-modular architectures in which components comprise software (Yoo *et al.*, 2010). Driven by the attributes of digitally stored information, software modules are often highly flexible and can be

implemented across design hierarchies (Yoo *et al.*, 2010). This hints at a tension in the literature on digital platforms as established logics in modularity are challenged in the context of platforms consisting of digital technology. Rather than designing, providing, and maintaining all functionality centrally and *a priori*, digital technology is malleable to changes after the fact (Kallinikos *et al.*, 2013b). Indeed, scholars have highlighted that the virtue of digital platforms is that their architecture allows the introduction of novelty through unprompted changes, often initiated by third-party developers (de Reuver *et al.*, 2018; Yoo *et al.*, 2010; Zittrain, 2006). While modularity offers a robust vocabulary for describing the design and organisation of digital technology, the platform literature hints at the opportunity for contributions through further investigations of the interaction between design and innovation on digital platforms. Constantinides *et al.* 2018 call for a “new mirroring hypothesis” in the context of digital platforms, pointing to the need of reflecting modular systems thinking with respects to digital platforms. Empirical research on digital platforms therefore has to strike a balance between leveraging and reflecting on the logics that modular systems thinking provides to platform scholars.

In summary, the ongoing discussion of digital platforms leaves room for substantial contributions for studies on digital platform growth and innovation. In particular, three broad lines for investigation stand out; (i), the detailed empirical study of digital platforms, (ii), the investigation of emergent phenomena such as generativity, (iii), the reflection on the intellectual traditions informing technology design and organisation.

How such contributions can be achieved will be addressed in the following chapter, which outlines the analytical approach adopted in this dissertation.

CHAPTER 3 – ANALYTICAL APPROACH

The preceding chapter reviewed literature on digital platforms and articulated the potential for further contributions. To address the potential, this dissertation adopts a computationally-intensive mixed-methods approach influenced by a modern pragmatist research tradition. Many of the phenomena of interest to scholars engaged with digital platform research are nascent in the sense that they are understudied and undertheorized, despite often being empirically apparent. This calls for an approach that can capture the phenomena of interest in a flexible manner. This section formulates and details the analytical approach that ties together the empirical studies presented in the remainder of the dissertation. The section argues that the dissertation's focus on the empirical study of digital platforms benefits from methods that can leverage contextual richness while also utilising available empirical material at scale. To that end, the approach adopted here aims at the detailed study of interaction with digital artifacts. The section proceeds in two stages. First, a brief reflection of computational approaches to IS research will be presented. Second, the use of mixed-methods designs will be embedded in a pragmatist research tradition.

3.1. COMPUTATIONAL RESEARCH IN IS

Like many phenomena of interest to information systems researchers, the dynamics on digital platforms are characterized by complex interactions between the social and the technical world (Lazer and Radford, 2017). And like many other technical systems, the interaction on digital platforms produces digital information continuously, simultaneously, and unequivocally from the interaction they afford with the user and other technology artifacts (Kallinikos, 2013). Recent theorisation efforts in the social sciences are permeated by an interest in the data collected from these systems, as well as the systems themselves (Alvarez, 2016; Cioffi-Revilla, 2010; Lazer *et al.*, 2009).

Social science research in general, and information systems research in particular, doubtlessly benefit from what others refer to as a “data revolution” (Kitchin, 2014a). Indeed, the growing accessibility and availability of data sources combined with impressive developments of computational tools for data collection and analysis has led to significant advances in social science research (Cioffi-Revilla, 2010; Lazer *et*

al., 2009; Salganik, 2017). By now, a number of novel and promising data sources are available for scholarly investigation. Proponents of computational methodologies claim that these novel data sources allow researchers to capture phenomena that were simply unobservable or even non-existing in the past (Agarwal, Gupta, and Kraut, 2008; Ahonen, 2015; Hedman, Srinivasan, and Lindgren, 2013). In the information systems field, many scholars seek to create an understanding of these changes in research practice that rests on computational tools without sacrificing the rigour needed in theory building.

While a comprehensive discussion of ‘computational social science’ (Cioffi-Revilla, 2010, 2014) is out of scope for this dissertation¹⁴, a brief reflection on computational research in IS is helpful. Information systems research has seen a rise in computational approaches and much of the debate concentrates on the purposefulness of computational tools, the utility of vast data repositories, and ultimately, the reflection on appropriating existing or developing new methodological approaches and the research paradigms inherent therein.

It is helpful to understand recent advances in IS research as something akin to a new paradigm¹⁵. Computational research promises to unearth details about the interaction of actors and technology artifacts at a level of detail that is arguably unprecedented. At the same time, vast data repositories enable the research of these interactions at a scale that was simply unavailable and unattainable for social scientists (see Salganik, 2017). Against this backdrop, scholars in the information systems field seek to find a way that aligns with the strong theory building tradition while utilising the promises that new approaches hold.

For example, by what has been labelled “zooming in and zooming out”, Gaskin *et al.*, (2014) have sought to conceptualize a research approach that leverages large scale

¹⁴ Note: the following sources offer more comprehensive discussions of the emergence, adequacy, and application of computational tools in the social sciences. They focus on a characterization of data sets, opportunities and pitfalls, as well as novel research designs for social scientists. These sources are referenced here as they informed the work undertaken in the dissertation but are not immediately relevant to this discussion: Salganik (2017), Marres (2015), Alvarez (2016), Goldberg (2015), Cioffi-Revilla (2010)

¹⁵ I use the term in the sense of Kuhn (1967;2000) but refrain from discussing paradigms and paradigm shifts in the context of the applicability of an epistemic stance; see Morgan (2007) and Kitchin (2014) for more detailed reflections on the Kuhnian sense of the term ‘paradigm’.

data sets while remaining focused enough to theorise the particularities of technology use in context. Using their language; zooming in provides opportunities for IS scholars to obtain a deep understanding of the interaction between social and technical spheres underlying various digital technology phenomena, zooming out of the particular often increases generalizability of the phenomenon under study. The integration of both approaches presents a critical frontier for information systems research. As a result, research in information systems undergoes a change as information systems scholars seek to reap the benefits of novel methods and data while maintaining the tenets of novel theory generation. As a result, a paradigm for theorization from computational approaches in the study of digital phenomena remains to be established.

Yet, computational research is increasingly recognised as a fruitful course of action for IS scholarship. For example, major scholarly outlets now embrace what is commonly termed as a “computational approach to IS research”¹⁶: For the purpose of this dissertation, a computational approach to IS research is defined as *the collection, analysis, and theorisation of process trace data to uncover engagement with the IS artifact* (compare Nambisan *et al.*, 2017; Rai, 2016).

Taking this definition as it’s point of departure, the following section formulates the research approach adopted in this dissertation by reflecting on three core aspects of the definition itself.

¹⁶ It is important to point out that despite the definition used here, no unified computational approach exists in IS research. Among the scholarly works that draw from computational methods in IS research, a separation akin to the qualitative-quantitative divide is discernible. This is not surprising. For instance, in *Chaos of Disciplines* (2001), Andrew Abbot discusses the long history of fractals in academe, i.e., recurrent patterns of distinctions among researchers that are observable on different levels of abstraction. To illustrate, he argues that while the entire field of Sociology can be divided in a quantitative and a qualitative research tradition, within either ‘camp’ one would find a similar division. As such, among scholars subscribing to the quantitative tradition, a spectrum would be discernible along which one could distinguish the degree to which individuals lean towards one or the other end of the spectrum. Recent debates in IS suggests similar ‘fractals’ exists among scholars interested in computational research.

3.1.1. The core activities in computational IS research

The use of computational methods spans the activity sequence of any archetypical research projects; (i) *data collection*, (ii) *data analysis*, and (iii) *theorisation*. The next section will present selected studies in each category. This stylizes the studies as belonging to one category but I acknowledge that most studies in IS transcend more or indeed all of the core activities outlined below.

(i) *Data collection*

Many studies leverage the availability of digital trace data for the study of identified or emergent phenomena using more or less established methods¹⁷. Often these computational data collection efforts provide insight into the handling and promises of unusual data structures. Lambrecht *et al.*, (2014), for example, collected search engine queries on mobile phones to study how dynamics known in marketing theory unfold when commercial activity is mediated by mobile technology (Lambrecht *et al.*, 2014). On the other hand Müller *et al.*, (2016), generate constructs from user reviews to associate formulations in the text with established constructs and instruments in the information systems literature to infer user satisfaction, for example (Müller *et al.*, 2016). Others use digital trace data to study how social processes, such as the formation of group and building of social ties, are reflected in the usage patterns on social media (e.g., Zeng and Wei, 2013).

(ii) *Data analysis*

Another group of studies focuses on the development or appropriation of novel ways of analysing computational data. This can be the case regardless of whether or not the data was collected by using computational tools or whether the data refers to a particularly “digital” phenomenon. These efforts are possibly most visible in the tools IS researchers employ for their analysis. Across levels of analyses that are of typical relevance for information systems research (e.g., individual, artifact, organisation, industry), the methodological toolbox of social science inquiry underwent a reappraisal. Gaskin *et al.*, (2014), for instance, formulate an approach to the analysis of sequence data from rich logs of traces from technology use

¹⁷ This section is chiefly concerned with research in the field of information systems (IS). For a summary of social science study designs beyond IS see e.g., Salganik (2017) or Alvarez (2016)

of individuals and teams. They thereby aspire to complement established analysis approaches such as variance and process based analysis by a sequential view. Eck and Uebernickel (2016) use network analysis and graph theory to infer dynamics across millions of software artifacts on software development data from GitHub (Eck and Uebernickel, 2016). Similarly, Johnson *et al.*, (2015) utilize natural language processing techniques to complement survey methods in the established quantitative deductive framework to infer how dynamics in online communities give rise to individuals with prominent positions (Johnson, Safadi, and Faraj, 2015).

(iii) *Theorisation*

Lastly, a number of studies seek to explicitly theorise distinctly computational phenomena. That is, phenomena that are of particular salience because they generated by interaction with digital technology. For instance, Lindberg *et al.*, (2016), study the coordination of open source communities by collecting and analysing data from the tools used to coordinate and discuss development work (Lindberg *et al.*, 2016). Similarly, Howison and Crowston (2014), combine ethnographic observations and the analysis of task traces left by users in the tools employed to coordinate work (Howison and Crowston, 2014). Yet another study by Aaltonen and Seiler (2016) uses data from Wikipedia to infer how the popularity of individual authors shapes editing activity in the emergent phenomenon of community content production. Other scholars embark on the theorisation of the supra-organisational activity that fuels recent innovation phenomena using both established econometric analysis (Um and Yoo, 2016) or more nascent techniques such as sequence and network analysis (Zhang *et al.*, 2014). Perhaps most encompassing, Berente *et al.*'s (2018), formulation of a grounded-theory inspired approach stands out. They conceptualise a research methodology that incorporates computational tools to collect and analyse data from the context in which technology use occurs in situ with the inductive iterations employed in grounded theorisation (Berente, Seidel, and Safadi, 2018).

3.1.2. Digital process trace data

The analysis of digital trace data is important to many pertinent information systems research topics. The term ‘trace data’ refers to all forms of digital information stemming from interaction with digital technology attainable through computational data collections (Agarwal *et al.*, 2008; Hedman *et al.*, 2013; Howison, Wiggins, and Crowston, 2011). This may include, but is not limited to, user logs from interaction with computing devices, source code data from version control systems such as GitHub, or archival records from mailing lists. As footprints of social activity enabled by digital technology, digital traces are especially valuable to the reconstruction of social interaction as well as the investigation of artifact-level data in great detail (Lazer and Radford, 2017). Not unlike archaeologists, IS researchers turn to the detectable residues of technology-enabled social interaction in complex interactions of digital artifacts surrounding our everyday activity (Yoo, 2010).

The trace data (Hedman *et al.*, 2013) produced from interaction with digital technology promise to fuel the study of artifact interact in great detail. For instance, constructing networks of interaction from source code revealed structural properties of a technology’s development at a scale that, until recently, was unfathomable (Baldwin, MacCormack, and Rusnak, 2014). Similarly, sequences of technical changes have been used to study interdependencies in hitherto latent development routines (Lindberg *et al.*, 2016). Other methods that have been added to the list of tools capable of dealing with trace data include ‘process mining’ (Pentland and Feldman, 2007), as well as text analysis (Debortoli *et al.*, 2016; Johnson *et al.*, 2015; Müller *et al.*, 2016).

Several aspects set digital trace data apart from data that most social scientists are trained to handle competently. As Hedman *et al.*, (2013) point out, the fact that trace data is of distinctly digital origin presents both opportunities and challenges for IS researchers. On the one hand, trace data helps reconstructing processes of interaction with digital artifacts at great detail (Agarwal *et al.*, 2008). This includes the opportunity to associate interactions to users, artifacts, locations, times, and context, all of which increase the level of detail for potential investigations (Hedman *et al.*, 2013). At the same time, most digital trace data sets were never generated for purposes of social science research. This introduces a number of issues that researcher need to reflect upon (see Hedman *et al.*, 2013 for a summary). First and foremost, digital traces are not created the same way as other empirical evidence in the social sciences. For

instance, survey researchers typically collect data through carefully designed studies employing established instruments. This introduces a series of concerns as issues such as construct validity have to be thoroughly established (Howison *et al.*, 2011; Salganik, 2017). For instance, Howison *et al.* (2011) dedicate an entire study to the discussion of validity of social networking methods when used to analyse data harvested from interaction with online information systems. Beyond established biases that play out similarly – or are indeed more exacerbated in digital traces – new forms of threats are also apparent. As Salganik (2017) points out, aspects of trace data sets, such as what he refers to as ‘drift’, introduce potential for error as the underlying system producing and disseminating trace data is subject to change by the operator of the system. While this is a matter of course for IS researchers (Kallinikos and Constantiou, 2015; Kitchin, 2014b), these threats are nonetheless novelties for most researchers. Since their existence introduce new sources of bias to analysis and inference, they merit reflection.

3.1.3. The IS artifact

Lastly, the definition of computational IS research emphasises the role of the IS artifact. This defines the focus of the investigation of information systems in terms of their functions and organisation. As Simon (1996) argues, what is meaningful about computers is not their material make-up and not the operations they can perform on the basic component level. What is meaningful, is the organisation of components to fulfil a task that is enabled by social actors interacting with technology (Simon, 1996). In line with this, the focus of computational IS research is the study of artifacts in terms of the interactions they render and enable. Miller and Page (2007) argue that the use of computations can approximate the dynamics of real-life technology interaction. The simulation of interaction can help in theory discovery as computations simulate the functions of an IS artifact that are pertinent to the phenomenon. With this in mind, they argue, computational approaches offer both the tools and the substance for theory discovery and building by focusing on the artifact at play (Miller and Page, 2007). Indeed, as the nexus of interaction between the social and the technical, the IS artifact has traditionally been of immense interest for the information systems field (e.g., (Grover and Lyytinen, 2015; Orlikowski and Iacono, 2001; Tilson *et al.*, 2010).

As outlined above, the phenomena of interest to this dissertation are subject to the underlying technology and thus inexorably tied to the artifact. With a reflection on the aspects of an established definition of computational research approach in IS, I here summarize the main characteristics of the research approach adopted throughout the dissertation.

First, the approach in the empirical work rests on the integration of different kinds of data obtained from artifact interaction all of which constitute attributes of digital traces (Hedman *et al.*, 2013). As such, the lion's share of empirical material that was collected for the purpose of this dissertation is represented in forms such as text, counts, or logs. Such data has been gathered using computational techniques such as interacting with standardised interfaces and databases (e.g., via web application programming interfaces; APIs), custom data collection methods (e.g., scraping web data); to using vast public data repositories (e.g., e-mail archives). The goal has always been to reconstruct the processes of interaction between different actors, or groups and/or with the platform as described in the introduction (Chapter 1).

Second, as is apparent from the reflections on digital trace data, many of the above data sources – while detailed reflections of technology interaction – are insufficient to guide a theory building process on their own accord. Working with digital trace data often necessitated making sense of patterns by corroborating initial interpretations. Therefore, the above data sources and the analysis techniques they often prescribed had to be complemented by other forms of evidence. Throughout the work that is presented in this dissertation, other data sources have thus been included ranging from informant interviews, over technical documentation, to archival data. With this in mind, the following section reflects on the use of mixed methods.

3.2. MIXED-METHODS DESIGNS

The above examples illustrate the growing acceptance of using computational analysis techniques in IS research. In order to be equipped to research the phenomena of interest in a domain that focuses on the interaction between the social and the technical, theory building from technically sophisticated approaches seems promising for IS research. Indeed, several recent studies in IS employ mixed methods designs, in the sense that they “gather both quantitative and qualitative data, integrate the two, and then draw interpretations based on the combined strengths of both sets of data to better understand research problems.” (Creswell, 2016; p. 2).

In many such study designs, the depth needed to detail social interaction with technical artifacts is complemented by computational analysis techniques. For instance, Vaast *et al.*, (2017) highlight how the qualitative analysis of large-scale datasets can be conducted in close interplay with computational-quantitative methods (Vaast *et al.*, 2017). In their study, Vaast and colleagues elaborate patterns of collective technology use that are motivated by a grounded inductive approach in an exploratory pre-study. Similarly, Sarker *et al.* (2018) follow a sequential research design in which they generate survey instruments from a preceding qualitative case study (Sarker, Ahuja, and Sarker, 2018). As many phenomena in IS are of contextually-defined character (Zachariadis, Scott, and Barrett, 2013), the combination of methods promises the generation of meaningful insights. Mixed methods thereby often provide the flexibility to investigate aspects of the phenomenon under study by exercising multimodality in techniques, data, and theoretical contributions (O’Halloran *et al.*, 2016).

Often, the appropriateness of different tools can only fully be evaluated once the researcher engages with the empirical material – and this evaluation can be subject to change over time (Kitchin, 2014). As a result, this dissertation does not subscribe exclusively to either qualitative or quantitative research traditions. Instead, techniques from one tradition (e.g., informant interviews) are augmented by the analysis of digital traces at scale (e.g., regression analysis). In line with the purpose of many mixed-methods designs (Venkatesh, Brown, and Bala, 2013; Venkatesh, Brown, and Sullivan, 2016) this supports the dissertation in scrutinising and contextualizing patterns derived from trace data. The promise of mixed-methods designs is the ability to explore understudied phenomena by corroborating, developing, or completing (Venkatesh *et al.*, 2013) perspectives on a phenomenon.

Creswell (2016) goes as far as to claim that “mixed methods might be seen as the first major social science research methodology in the 20th and now 21st century to fully utilize the digital capabilities to advance it.” And indeed, the scholarly community using mixed-methods-methods designs embraces the developments that can be summarised as computational research (e.g., Castro *et al.*, 2010; Fielding, 2012; O’Halloran *et al.*, 2016)

3.2.1. Pragmatist Influence

In this sense, the efficacy of multiple research traditions under the umbrella of the same research project is herein viewed in a pragmatist sense. That is, by asking which paradigm is more suited to explain the phenomenon of interest, a pragmatist answer to fundamental question of the scientific process recognises that no methodological tradition in isolation can exhaustively capture and explain digital platform phenomena. The answer to the question of what is there to know about platforms and how can it be known simply requires to incorporate qualitative as well as quantitative methods (Morgan, 2007; Venkatesh *et al.*, 2013).

The following section briefly reflects on a recent form of pragmatist analysis¹⁸. As the modern form that influenced the work in this dissertation, it is distinct from the strong ontological assumptions in pragmatism put forth elsewhere (Farjoun, Ansell, and Boin, 2015). Instead, current portrayals of pragmatist analysis in IS depart from tied interlocking of ontology, epistemology, and methodology as the sole paradigm to conceptualizing social science research designs (Morgan, 2007; Venkatesh *et al.*, 2013).

Central to the interest and the contribution of pragmatist studies is human action (Lorino, 2018). Pertinent pragmatist questions are thus; ‘what is being done, by whom, and what for?’ (Goldkuhl, 2004). At a general level, contemporary forms of pragmatism refer to the practice of producing knowledge about a phenomenon in a way that is useful to those who are interested in the phenomenon (Lorino, 2018; Wicks and Freeman, 1998). The aspired outcome from pragmatist analysis is hence validated

¹⁸ For detailed discussions of the historical perspective on the philosophical tradition of pragmatism based on Pierce, Dewey, and James see (i) Lorino, 2018, (ii) Paavola 2015, (iii) Farjoun *et al.* 2015, as well as The Stanford Encyclopaedia of Philosophy (<https://plato.stanford.edu/>)

by a focus on process, action, and consequences. Subsequently, the value of a contribution is determined by the utility and efficacy of claims about a world that is reflected in human action and as such is “out there but not objective” (Wicks and Freeman, 1998; p.126). The analytical approach in this dissertation is influenced by this view on pragmatism in social research. In particular, several aspects of a modern pragmatist stance present alignment with a computationally-intensive mixed-methods approach to IS research. In essence, a pragmatist perspective provides an intellectual foundation for inquiries concerned with the study of human action (Farjoun *et al.*, 2015; Lorino, 2018).

On the one hand, the influence of pragmatism on this dissertation presents a mean as pragmatism embraces a plurality of methods. In a pragmatist view, a contribution can be achieved through a variety of activities. As long as an understanding can be achieved that enables the advance of shared knowledge among the members of the “community of inquiry” (Constantinides, Chiasson, and Introna, 2012; Lorino, 2018). Of course this does not equate to an ill-reflected “*anything goes*” approach to research designs (Wicks and Freeman, 1998). Instead, the tenet for combining diverse data sources and appropriating methods is the usefulness of the knowledge contribution to a community of inquiry. Indeed, the community for which a contribution is useful is not necessarily characterised by its shared practices but by shared interest and concerns (Lorino, 2018). When it comes to mixed methodologies, this implies that methods can be flexibly applied to corroborate insights if that yields usefulness for the addressed audience – be that scholarly or otherwise. Scholars therefore often exercise a pragmatic freedom by combining methods in order to make sense of patterns and relationships as they appear reflected in the empirical material. Pragmatist analysis draws distinctions between methods and forms of evidence only in so far as an approach ought to yield utility (Wicks and Freeman, 1998). Therefore, it is not only pragmatic to combine qualitative and quantitative methods since that seems an appropriate way to theorizing emergent phenomena. Also, it is in the interest of pragmatist analysts to give information that is “useful in the sense of helping people to better cope with the world or to create better organizations” (Wicks and Freeman, 1998; p.129). What is of use is thereby irrespective of the form of evidence.

A pragmatist streak in information systems research increasingly recognises the efficacy of combining qualitative and quantitative methods as the boundaries between

research traditions blur (Morgan, 2007). The need to combine methods is exacerbated by the aforementioned push in IS research to conceptualise approaches to computational data of novel kind and unusual structure. Giving priority to research questions over epistemological stance, pragmatism is thus inclusive as opposed to exclusive of different methodologies (Goldkuhl, 2012; Morgan, 2007; Yin, 2015). This does not only include bridging quantitative and qualitative epistemologies, but pragmatism's "third way" (Yin, 2015) also speaks to the character of computational research in IS, which in the understanding outlined above, is pluralistic in methodology as well as research practice.

On the other hand, pragmatism serves the end of creating knowledge that is useful and actionable. Pragmatism aspires to contribute insights that are tangible. The pragmatist contribution can therefore be seen as practice in a broad sense. As such, pragmatist studies in IS are characterized by any number and combination of the following foci (compare Goldkuhl, 2004);

- (i) pragmatist function, that is to create useful knowledge,
- (ii) pragmatist reference, that is studying social actions, or,
- (iii) pragmatist method, that is to engage in or prescribe action.

It is important to point out that pragmatist contributions are not necessarily interventions. That is, pragmatist studies do not require interventionist designs such as action research (e.g., Baskerville and Wood-Harper, 1998) or design science research (e.g., Hevner *et al.*, 2004). Instead, pragmatist contributions include actionable insights such as recommendations for actions, or the prescription of action as an equally acceptable contribution to knowledge (see Goldkuhl, 2011; Wicks and Freeman, 1998). This further aligns a pragmatist epistemology with computational methods as such contributions can be derived through a variety of representations of social interaction using computational techniques (Miller and Page, 2007). Useful contributions may thus include predictions, recommendations, simulations, or simply detailed empirical observations. As Miller and Page (2007) argue, employing computational analysis serves theory generation and discovery as such tools are able to simulate complex interaction rendered by technology – whether in fact or in hypotheticals (Miller and Page 2007).

In the pragmatist tradition, the main mode of reasoning is abductive (Lorino, 2018; Kitchin, 2014). That is, inductive and deductive approaches are combined either explicitly or implicitly¹⁹. This aims to advance the understanding of a phenomenon through iteration between data, theory, and analysis. Abduction therefore “seeks to generate hypotheses and insights “born from the data” rather than “born from the theory” (Kitchin 2014b; p. 138). True to the pragmatist tradition of generating useful knowledge, is the understanding of abduction as reasoning about “first suggestions” (Peirce, 1992, cited in Bamberger, 2018). As such, abduction is an inferential mode of analysis in that it offers the interpretation of knowledge by converging to a theory of best fit (Ketokivi and Mantere, 2010). This often follows approaches that either challenge existing theory or build first theoretical conjectures of plausible explanations (Bamberger, 2018).

Kitchin (2014) describes how the abductive mode of reasoning extends to computational methods and the pluralism exercised therein. He argues that the decisions determining which tools to use for data collection, analysis, and report follow the same logic as abductive theorising. That is, methods are chosen according to what is known as well as what makes sense in a specific situation (Kitchin, 2014). In so doing, pragmatist analysis typically moves back and forth between induction and deduction within and across phases of a research project. Typically, this form of analysis unfolds sequentially, and inductive propositions precede deductive confirmation (Kitchin, 2014). This is mainly achieved by a continuous iteration between empirical material, extant theory, and explorative analysis. As such the pragmatist analyst first converts observations into working hypotheses, then assesses those conjectures through pragmatist values and re-iterates exploratory analysis steps, before embarking on theorising (Morgan, 2007; Paavola, 2015; Venkatesh *et al.*, 2013).

Pragmatist approaches are inherently theoretical. Ultimately the pluralism presented towards methods and data in pragmatist analysis extends to the process of theorisation. While no theory is given preference over another. The maxim under which theoretical

¹⁹ see Paavola (2015) for a comprehensive discussion about the origin of abductive reasoning in Pierce’s and Dewey’s writings.

contributions are being evaluated is usefulness in describing the phenomenon, and interest to the community of inquiry. This underlines the interest of pragmatist approaches to contribute theoretical insights not for the theory’s sake, but rather for the, sake of useful knowledge.

The studies presented in this dissertation utilize the approach outlined above along the following lines. First, the studies draw from a variety of data sources and integrate data of different kind. Second, in order to leverage insight from the data multiple analysis techniques are combined within the respective studies. Third, the studies are characterized by an abductive approach to theorisation in that phenomenon, available evidence, and theoretical conjectures were derived through continuous iteration between inductive and deductive reasoning about emergent insights.

Finally, the contributions aspired in each study focus on an empirical phenomenon of salient interest to scholars in the digital platform space. In order to contribute to the debates pertinent to that community, the methods employed are of importance only in so far as they help to arrive at contributions that presents useful knowledge. This is not to say that every study in this dissertation employs a full mixed-methods design aimed at deriving recommendations for practitioners. This is to say that in order to leverage insight in the presented studies, a variety of data sources were integrated and decisions made to analyse the evidence were made flexibly in order to answer a research question of interest to the scholarly platform community. As is summarised in table 3.1 below, this was often achieved by combining methods and data sources.

The table gives an overview over the studies presented as part of this dissertation, the data sources, analysis techniques used, and the pragmatist aspects in each study.

TABLE 3.1. Analytical Approach in the Three Empirical Studies				
Study	Focus	Data sources	Analysis Techniques	Pragmatist Influence
#1	Endorsements by platform operator and the impact on generativity	1. Community Forum 2. Tagged geo-data	1. Qualitative content analysis 2. Quantitative analysis in a generalized linear	1. Method pluralism 2. Focus on action: ‘endorsements’ 3. Contribution: recommendations to platform operators

			modelling framework	
#2	Platform module designs and the effect for the pace of change	1. Source code of platform core modules	1. Coding using technical documentation material 2. Longitudinal panel analysis using mixed effect estimators	1. Focus on action: 'module designs' 2. Contribution: Strategic option for platform operators recommends course of action
#3	Coordination among platform participants	1. E-mail archive 2. Source code development logs 3. Informant Interviews (secondary and primary)	1. Quantitative analysis 2. Content Analysis 3. Network Analysis	1. Method Pluralism 2. Focus on action: 'coordination' / 'contribution'

Typically, the approach followed in this dissertation was characterised by three steps - not necessarily in sequential order.

First, finding groupings in data either through naïve imposition of groups and subsequent confirmation, or through induction of groups for instance by using computational techniques such as unsupervised clustering algorithms. Second, groups in the data were then examined with the aim of establishing meaningful differences between them. This step can include either a theoretically-informed reasoning about groups or a quantitative test of significance of a found or claimed difference between observations. Lastly, groups in the data and the attributes that differentiate them were expressed in a formal relationship. This happens either by surmising a causal relationship derived from theory or by modelling a causality quantitatively by ruling out alternative explanations in form of a statistical model. In line with the pragmatist tradition, the theorisation process then involved the evaluation and suggestion of an interpretation. This was then reported in a theoretical frame deemed most adequate to illustrate a plausible explanation rooted in the empirical work.

The following chapters four through six each present an empirical study.

CHAPTER 4 – THE IMPACT OF ENDORSEMENTS ON PLATFORM GENERATIVITY

ABSTRACT

Platform generativity – the generation of new outputs without direct input by the platform operator – is highly desirable for platform operators. However, not all generativity is equally desirable. Platform operators may therefore have strategic motives to encourage contributions by platform participants in areas where generativity is needed. This is done through endorsements, that is, actions that signal desirable interaction on the platform. In this paper, we investigate how and why endorsements influence generativity. To this end, we adopt a mixed-methods design of the geo-data platform OpenStreetMap. First, we conduct an in-depth content analysis of its discussion forum to inductively identify strategic motives for endorsements. Second, we then formulate hypotheses to test the impact of these endorsement motives on platform generativity using an original data set of tagged geo-data from OpenStreetMap. Our research offers contributions to the platform literature by (1) identifying motives for endorsements that platform operators enact to signal areas where generativity is desired, and (2) testing the impact of endorsements with different strategic motives for platform generativity.

A version of this chapter has been submitted for publication and is currently under review as:

Hukal, P., Henfridsson, O., Shaikh, M., Parker, G., (2018), THE IMPACT OF ENDORSEMENTS ON PLATFORM GENERATIVITY

THE IMPACT OF ENDORSEMENTS ON PLATFORM GENERATIVITY

4.1. INTRODUCTION

There is little doubt that the success of digital platforms largely depends on the input of its participants. To this end, platform operators typically make efforts to stimulate input from platform participants. Ideally, such efforts foster generativity, that is, the generation of new outputs without direct input from the platform operator (Wareham *et al.*, 2014; Zittrain, 2006). Generativity increases the value of a platform as it reflects continued use and re-use of what the platform offers (see e.g., Tilson *et al.*, 2010; Wareham *et al.*, 2014).

However, not all generativity is unequivocally desirable. The platform operator might therefore have strategic motives to increase generativity in one particular area of platform activity, but not in others. Enacting such motives, one significant challenge for the platform operator wishing to direct generativity is the information asymmetry between the platform and its participants (Ho and Rai, 2017). Platform participants cannot know where platform generativity is desirable without any form of communication. To this end, the platform operator can signal its intentions to platform participants. For instance, the platform may bestow special status to preferred developers (Ho and Rai, 2017), sanction a new use case (Förderer *et al.*, 2018), or formulate code of conducts (Karhu *et al.*, 2018), or announce future plans (Parker *et al.*, 2017). We refer to these signals as endorsements. The platform operator is a sender of information that reduces the information asymmetry, while platform participants are receivers interpreting the signals communicated by the platform (Connelly *et al.*, 2011).

In this paper, we examine endorsements as actions used by platform operators to increase generativity. Anecdotal evidence suggests that endorsements have different strategic motives²⁰ that resonate with strategies put forth in the platform literature. While prior literature documents examples of endorsements (Hanseth and Lyytinen, 2010; Ho and Rai, 2017; Parker *et al.*, 2017), little has been done to empirically investigate the impact of endorsements on platform generativity and whether the

²⁰ For instance, platforms might have “white spaces” where more input is wanted. One of the co-authors have collaborated with SAP with a focus on their platform developer program. In this work, SAP managers referred to areas where they wanted third-party input as “white spaces”.

strategic motive behind an endorsement matters. Yet, this is increasingly important for anyone responsible for managing a platform and its future growth. Our paper therefore deals with the following research question: *How and why do endorsements influence generativity in digital platforms?*

To address this research question, we adopt a multi-method design (Venkatesh *et al.*, 2013) to analyse endorsements on the geo-data platform OpenStreetMap. Powering services such as Craigslist and Foursquare, the generativity of OpenStreetMap is essential to continue to serve as a viable alternative to giants such as Google Maps. We collected 5 years (May 2009 to December 2014) of data related to OpenStreetMap and its endorsements to stimulate platform generativity. First, we conducted an in-depth content analysis of the OpenStreetMap discussion forum to inductively identify different strategic motives for endorsements. This analysis yielded four strategic endorsement motives: commit to new market; accommodate third-party request; balance market demand; ratify emergent use. Second, we formulated hypotheses to test the impact of these four endorsement motives on platform generativity. For this confirmatory quantitative analysis, we use an original data set of tagged geo-data from OpenStreetMap.

Our research offers a contribution to the platform literature by (1) identifying motives for endorsements that platform operators can use to signal areas where generativity is desired, and (2) testing the impact of those endorsements for platform generativity.

4.2. PRIOR LITERATURE AND CONCEPTUAL BASIS

Even though generativity as a notion was used as far back as in 1985 (Lyytinen, 1985) it has only recently received significant attention in the IS literature. A dictionary definition of generativity reads: it “generates, produces, or gives rise to something, or has the power or ability to do so; productive, creative; originating, causative”²¹. Most prior literature dealing with generativity broadly fits this definition. Generativity is typically seen as a self-reinforcing process triggered by interaction between the platform and its participants in a way that leads to the creation of novel products and services (Yoo *et al.*, 2010, Parker *et al.*, 2017). The derivative forms of interactions are often unanticipated by the platform operator, hence realising use cases beyond the

²¹ <http://www.oed.com/view/Entry/77523>; accessed 03-06-2016

initially intended design of the platform (Tilson *et al.*, 2010; Wareham *et al.*, 2014; Yoo *et al.*, 2012, 2010).

As indicated in the introduction section, we define generativity as the generation of new outputs without direct input from the platform operator (cf. Wareham *et al.*, 2014; Zittrain, 2006). First of all, depending on the type of digital platform, the outputs of generativity can materialise in terms of transactions (e.g., in a marketplace), applications (e.g., in smartphone platforms such as iOS and Android), or contents (e.g., on data platforms such as OpenStreetMap).

Second, generativity typically increases the possible number and diversity of platform outputs (cf. de Reuver *et al.*, 2018). On an abstract level, platform generativity is a function of simultaneously increasing scale and scope of a digital platform²². Platform scale refers to the self-reinforcing processes of a growing number of outputs as a function of the installed base (Parker and Van Alstyne, 2005). Scale also alludes to an increased use and re-use of platform resources enabled in a platform's architecture. As indicated in Parker *et al.* (2016), the number of interactions involving the platform and its participants are fundamental to the value it offers (Parker, Van Alstyne, and Choudary, 2016). The number of outputs, whether they consist of transactions, applications, or contents are equivalent to touch-points on the platform.

In addition, outputs from generativity can also increase a platform's scope (Gawer, 2014). We refer to platform scope as the range of outputs produced by the platform and platform participants (Gawer and Cusumano, 2013). In this way, scope relates the number of actors, products, or transactions to their diversity. The platform scope increases with the number of alternative outputs that are enabled by touch points of different kinds.

Generativity typically extends platform scale and scope simultaneously by stimulating “exponentially growing ecosystems” (see de Reuver *et al.*, 2018; p.2). This view is mirrored in the different perspectives on platforms (see Gawer, 2014). In studies on platforms as marketplaces, scale and scope allude to the importance of direct and

²² for instance, Lyytinen *et al.* (2018) define generativity as “the capability and related mechanisms for unbounded growth in scale and diversity of the functions and embeddedness of the [digital technology] infrastructure.”

indirect network effects among platform actors (Gawer, 2014; McIntyre and Srinivasan, 2017). Here, scale refers to the idea that the value of a platform is proportional to the number of actors on the platform (Evans, 2009; McIntyre and Srinivasan, 2017). On the other hand, scope implies offering specialised output, that is, enabling exchanges between actors of different kind. Bundling resources so as to realise complementarities among actors is crucial for platform success (Eisenmann *et al.*, 2011). Parker *et al.* (2017), for instance formalise the benefits for platforms if network externalities successfully cross distinct user groups (Parker *et al.*, 2017). This notion is also echoed by studies with an interest in platforms as modular technologies. Here platform generativity is seen as realising scale and scope by facilitating interaction through architectural interface design (e.g., Yoo *et al.*, 2010; Eaton *et al.*, 2015). Scale addresses an increased use and re-use of platform resources, while scope refers to increasing the range of possible interactions enabled in a platform's architecture. Hence, scale arises when more and more actors draw from the resources offered on the platform whereas scope increases when the platform architecture provides many alternative ways of interacting with the platform. The combination of both is what is considered a driver of "serendipitous interactions that brings forth novel and unanticipated uses" (Yoo, 2012).

Finally, the generativity definition also stresses that the generation of output is accomplished without direct input of the platform operator (cf. Wareham *et al.*, 2014; Zittrain, 2006). Consider that an arm's-length relationship with platform participants is central to platform governance (Ghazawneh and Henfridsson, 2013), since the platform operator cannot engage in each one of the many new interactions that it offers to platform participants. It therefore exists an information asymmetry. Without direct input, platform participants cannot possibly be sure about the platform's intentions. Yet, in cases where generativity is desired, outputs by the platform participant are imperative. Indeed, a platform has significant motives to direct generativity to particular areas of the platform in order to increase platform scale and scope. As others have pointed out, by selectively stimulating both the number and diversity of actors, products, or transactions on a platform, platform operators manage and sustain platform growth (Parker *et al.*, 2017).

4.2.1. Endorsements and Generativity

The question of how to trigger generativity on platforms is of salient interest in the platform literature. The answer has typically involved some kind of governance mechanism aimed at managing openness of and access to the platform in order to increase scale and scope. In the logic of platforms as markets, for example, the desired network effects are triggered if marketplace access is discounted for one or more party involved in multi-sided transactions (e.g., Boudreau, 2010; Evans, 2009). Similarly, component designs have been suggested to facilitate new interaction with platform core resources. In what has been called ‘boundary resources’ platform operators can deploy components at the platform boundary to instantiate arm’s-length relationships with external participants on the platform (Ghazawneh and Henfridsson, 2013; Eaton *et al.*, 2015). In acts of “resourcing and securing” (Ghazawneh and Henfridsson, 2013), platform operators can thus enable or constrain interaction across the platform boundary.

We here suggest that there is another mechanism available to platform operators: signalling its intentions. Past studies assume generativity as a function of purposeful design (Hanseth and Lyytinen, 2010). In this regard, platform operators are encouraged to identify mechanisms that are capable of steering behaviour among participants (Wareham *et al.*, 2014). A prominent tool is endorsing a standard to increase the ease of adoption (Tiwana *et al.*, 2010). This would decrease the effort needed to engage with a platform, and endorsements therefore help directing the platform by means of a signalling effect (Evans and Schmalensee, 2016; Terlaak and King, 2006)

Prior platform literature has highlighted the need to orchestrate interactions on platforms (Benlian, Hilbert, and Hess, 2015; Wareham *et al.*, 2014). However, the role of signals in the form of endorsements remains underdeveloped. Clearly, endorsements should signal how the platform operator envisions the interaction among participants. For instance, endorsements signal one’s confidence in the endorsed entity by attracting attention to it (Boudreau and Jeppesen, 2015; Evans and Schmalensee, 2016). Signals can also make future intentions of the platform operator known which in turn enables participants to reliably engage with the platform and by extension with each other (Chintakananda and McIntyre, 2014; McIntyre and Srinivasan, 2017). Similarly, endorsements stimulate the potential for innovative

recombination of platform resources by signalling which kind of interacting with platform resources is desired (Ho and Rai, 2017).

However, further work is needed into how and why endorsements by the platform operator act as signals that stimulate generativity. We develop an understanding of generativity of digital platforms by focusing on the expansion of platforms in scale and scope through endorsements. This will complement the literature's profound interest in generativity (e.g. Eaton *et al.*, 2015; Svahn *et al.*, 2017; Tilson *et al.*, 2010; Wareham *et al.*, 2014) with a more focused look at how and why platform generativity can be supported through signals such as endorsements.

4.3. RESEARCH DESIGN

In this section we detail the empirical context of the geo-data platform OpenStreetMap and present our analytical approach.

4.3.1. Empirical Context

OpenStreetMap (OSM) is an open source platform that provides geo-spatial data on a global scale for free²³. In operation since 2004, it is considered the world's largest geo-data project on the web. Dubbed '*the Wikipedia of maps*'²⁴ it has over 4.5 million registered users of which more than 1.0 million contribute content to the map each month²⁵. All of OSM's geo-spatial data is governed by an *Open Data Commons Open Database License*²⁶, and is widely used in configuration of various web services: Over 350 external parties – commercial and non-commercial – are known to be using OSM data for their services, including popular providers such as Apple, Wikipedia, Foursquare, and Craigslist²⁷.

On OSM, 'mappers' typically contribute data by uploading GPS points to the OSM database called 'traces'. The GPS data is the foundation of every object in the database

²³ For a comprehensive introduction to the OpenStreetMap project see; Ramm, Topf, and Chilton (2010), *OpenStreetMap Using and Enhancing the Free Map of the World*, Cambridge: UIT Cambridge

²⁴ <http://www.theguardian.com/theobserver/2012/feb/18/openstreetmap-world-map-radicals>; accessed 03-02-16

²⁵ For additional statistics see <http://wiki.openstreetmap.org/wiki/Stats>; accessed 21-06-18

²⁶ In its current version 1.0 is a "ShareAlike license"; in accordance with the Open Data Commons (ODC) framework this allows free data use and re-use as long as the using party grants the same rights to re-users; details at <http://opendatacommons.org/licenses/odbl/>; accessed 06-07-15

²⁷ See http://wiki.openstreetmap.org/wiki/List_of_OSM-based_services; accessed 03-02-16

(Mooney and Corcoran, 2012). Data objects are grouped into points, lines, or polygons to represent all sorts of geographical features. Uploaded geo location data is subsequently annotated with semantic information through free-text labels – so called tags. OSM tags complement the data model and take the form of *key = value* to describe real-world objects. Tags are constructed such that the key term indicates the general category of an object, whereas the value term qualifies attributes of a key. For example, the key, *highway=*, is a high-level term to denote OSM objects representing streets and associated objects. The value component of OSM tags helps qualifying OSM elements by specific kind e.g. *highway=motorway*. For illustration, popular *highway* combinations are displayed in figure 1. Given the logic of combining geographic and thematic dimensions, the data model can flexibly represent real-world objects of high complexity. For instance, tags can represent everything from post boxes, trees, shops, streets, to building complexes, administrative boundaries, or transport routes.

Figure 4.1. OSM Tag Combinations

Tag: highway=residential
Tag: highway=track
Tag: highway=service
Tag: highway=unclassified
Tag: highway=footway
Tag: highway=path
Tag: highway=tertiary
Tag: highway=secondary
Tag: highway=bus_stop
Tag: highway=crossing

Note: The ten most used tag combinations for the key highway= in Europe as of Dec 1st 2014

OSM data objects can carry an infinite amount and combinations of tags to describe objects in great detail. Tags on OSM objects start off deliberately arbitrary: any term is allowed to describe a node, way or relation as long as it is "accurate and current" (Ramm *et al.*, 2010; Mooney and Corcoran, 2012). Any registered user and users of third party applications can add, delete, or edit any number and any kind of OSM tags on any OSM object at any point in time. Given the number of users, this results in continuous iterations of edits of OSM objects (Arsanjani *et al.*, 2015; Mooney and Corcoran, 2014). Users, mappers, and developers constantly update tags to reflect real world changes to natural or man-made objects such as the construction of a building.

For instance, tags might be added in order to further qualify objects. An OSM data object might be thematically described as a generic street first, later specified by street name, before being tagged with details such as speed limit, lighting, name, postcodes etc. (see figure 4.2).

Figure 4.2. Tag Edits on an OSM highway Object over Time²⁸

<pre><way id="227570282" timestamp="2013-06- 26T20:32:56Z" version="1" <tag k="highway" v="secondary"/> <tag k="lit" v="yes"/> <tag k="name" v="Abbey Road"/> <tag k="ref" v="B507"/> <tag k="sidewalk" v="right"/> </way></pre>	<pre><way id="227570282" timestamp="2013-06- 26T20:37:14Z" version="2" <tag k="highway" v="secondary"/> <tag k="lit" v="yes"/> <tag k="name" v="Abbey Road"/> <tag k="postal_code" v="NW8"/> <tag k="ref" v="B507"/> <tag k="sidewalk" v="right"/> </way></pre>	<pre><way id="227570282" timestamp="2013-12- 08T23:00:32Z" version="3" <tag k="highway" v="secondary"/> <tag k="lit" v="yes"/> <tag k="maxspeed" v="20 mph"/> <tag k="name" v="Abbey Road"/> <tag k="postal_code" v="NW8"/> <tag k="ref" v="B507"/> <tag k="sidewalk" v="right"/> </way></pre>
---	--	---

4.3.2. The role of metadata for generativity

The context is well suited to study generative change and endorsements on platforms. Geo-spatial data is widely interwoven with the fabric of a number of digital products and services (Parsons, 2013; Varian, 2014). The accessibility and availability of OSM data invites heterogeneous actors to create, combine, and reuse geo-data in order to configure novel products (Nambisan *et al.*, 2017). OSM data forms an important building block for digital services without a need for third-parties to be associated with the OSM organisation, or possess demanding geo-spatial information systems capabilities (Parsons, 2013; Yoo, 2013). Through the interaction with OSM data, the database now enables use cases beyond initial design, central control, and deliberate plans by the operators of OSM (Ramm *et al.*, 2010, Wareham *et al.*, 2014).

In the context of OSM, tag endorsements are an important lever for triggering generativity on the platform. Viewing generative changes as simultaneous increases of scale and scope translates into two attributes of the geo-data that an endorsement can affect. First, growing tag usage, that is the number of objects that is described by a key=value combination. Tag usage captures the notion of the platform growing in scale as more and more objects are described by a specific tag. However, growth in tag usage is but one dimension of our understanding of generativity. A second important aspect is the level of detail with which objects in the database are described

²⁸ Username, object and upload history are omitted for illustration; last accessed 11-06-2016

(that is, the number of metadata tags on a data object). On OSM an object in the database (a plain geo-location) can be described by any number and any combination of tags. The more detailed the descriptions of a data object, the more diverse the ways in which it can be used. An OSM object with a high number of tags indicates a rich and detailed thematic description of an otherwise plain location coordinate as shown by the panel in figure 2. More metadata enables using and re-using the data object in a variety of ways.

For an endorsement to be a potent signal for where generative change is desired it needs to increase both the scale and the scope of a platform. In the context of OpenStreetMap, we surmise that generative change is present if endorsed tags grow in usage (scale) and if endorsed tags increase details in the metadata (scope).

4.3.3. Analytical Approach

We aim at generating rich insights into how generative change on platforms can be simulated through endorsements by platform operators. We therefore follow a mixed-method approach (Venkatesh *et al.*, 2013). Our methodology can be classified as “mixed-methods multistrand” (Venkatesh *et al.*, 2016), in a “sequential exploratory design” (Creswell, 2016) where we derive categories and relationships from an exploratory qualitative analysis, which are then tested in a confirmatory quantitative framework. This approach presents the opportunity to distil insights from rich data sources while leveraging the full scale of the available empirical material.

Given the absence of a clear frame of reference for an empirical study of generativity, a mixed-methods design provides the flexibility that is crucial for exploration and initial theorisation of an understudied phenomenon. Mixed methods approaches have recently gained popularity in the information systems discipline. Very similar to our approach here, Sarker *et al.* (2018), for instance, hypothesise from a qualitative case study and test a derivative research model with survey data. Berente *et al.* (2018) advocate a computationally-intensive paradigm to information systems research that leverages grounded approaches in conjunction with computational techniques for data collection and analysis. The purpose for information systems researchers to implement mixed method approaches is the promise of theorising from multifaceted empirical evidence of social interaction with digital technology. In this sense, our analysis is

grounded in the empirical setting, developing theory in close dialogue between extant theory and diverse forms of empirical evidence.

Specifically, we apply a multistrand mixed methods design with a sequential implementation (Venkatesh *et al.*, 2016). We aim at connecting theory and empirical material by deriving ‘bridge assumptions’ (Kelle, 2015) to guide the analysis. As such, we generate theoretical propositions through a qualitative content analysis and subsequently evaluate identified categories and their interrelationships through hypothesis tests. Through a quantitative-dominant approach the quantitative analysis is motivated and guided by prior qualitative insights (Venkatesh *et al.*, 2016; Creswell, 2016). This sequence of steps underscores the study’s goal to advance our understanding of the empirical character of generativity through exploratory and confirmatory steps, culminating in a pragmatic analysis of the constructs and their interrelationships thus derived (cf. Morgan, 2007).

We proceed with the analysis in two phases. We first induct different motives of endorsements on the OSM platform. We subsequently test our hypotheses using quantitative analysis of the collected tagged geo-data in our research database.

4.4. ANALYSIS STAGE 1: CONTENT ANALYSIS AND HYPOTHESES

As an inductive process, this part of the analysis was marked by multiple iterations between the empirical evidence and prior work in the platform literature. In our case, the induction of testable propositions was especially valuable as it helped us to generate insights from patterns and associations in the data in close interplay with the lexical vocabulary of platform scholarship (cf. Berente *et al.*, 2018). From our in-depth qualitative content analysis, we hypothesise four motives for endorsements that are expected to trigger generativity on platforms. The generated hypotheses are not only of high theoretical interest but are thoroughly grounded in “actual settings and processes” (Vaast and Walsham, 2013).

4.4.1. Data

We studied the strategic motives behind platform endorsements by collecting data from the OpenStreetMap wiki forum. We sampled ten general purpose tags from the platform in a purposive selection (Yin, 2015).²⁹ As the starting point for our mixed methods design, we aimed at maximum information richness while ensuring analytical generalisability from our data (Creswell, 2016; Rapley, 2015). With the aim to reflect the number and diversity of outputs on the platform, we therefore sampled tags according to criterion similarity on usage (number of objects), variety (number of key=value pairs), and purpose (useful for mapping applications) (see Yin, 2015; Kelle, 2015). The selected tags are foundational to mapping in general and thus used widely. The tags were also of high relevance for platform participants. For example, OSM objects with *highway* tags are utilised in navigation applications using tags such as kind, surface, and quality of streets. These tags enable routing computations, traffic estimations, or public transportation applications³⁰. We provide a detailed overview over the sampled tags in the appendix.

The OSM community exercises no direct control over which tags are added to the platform. Yet, through processes of negotiation and discussion via various channels (Mooney and Corcoran, 2012), the core members of the community encourage the use

²⁹ The ten tags were: access, amenity, building, bridge, highway, landuse, leisure, service, shop, and railway; see appendix for details.

³⁰ As an indication, as of May 2016, 24 external projects were registered as using highway tags for the rendering of their services. This is based on self-reporting here <http://taginfo.openstreetmap.org/keys/highway#projects>; accessed 08-05-2016

of certain agreed-upon tag combinations. Once a tag is documented (i.e. being discussed publicly), it is assigned a status that is shown on the OSM wiki forum. The OSM wiki represents the main source of information for community members (Davidovic *et al.*, 2016; Ramm *et al.*, 2010). Thus the endorsement of a tag has project-wide consequences as the central stimulus is intended to communicate an otherwise unobservable quality to platform participants (see Ho and Rai, 2017). We therefore operationalise an *endorsement* as whether or not a given tag has been accepted by the community as indicated on the central wiki environment³¹.

Despite OSM's community-based collaboration (Budhathoki and Haythornthwaite, 2013), tag endorsements are akin to actions by a platform operator. We know from our work with the case that only a small, recurring fraction of the overall user base engages in the tag proposal processes. These community members are often prolific mappers and devoted to the platform. The decisions made by this very small number of community members guides the direction of the entire platform and influences how vast numbers of participants interact with it. The role of an endorsement as a signal about how and where to engage with the platform is therefore noteworthy because it exercises control over the direction of the platform. The chosen direction is explicated by an endorsement. In the case of OpenStreetMap, the larger community of participants accepts the direction provided by the central stimulus as this embeds decisions into technology (Lessig, 1999; Zittrain, 2006, 2008) thus making the endorsement objective and unproblematic – akin to a standard (Hanseth and Lyytinen, 2010).

Tags are typically endorsed through a proposal process initiated by a platform community member. Endorsement proposals are dedicated web pages in the community wiki focusing on one tag. They outline the rationale of why a tag should be endorsed, give examples, and reference how the potential tag aligns with the existing tagging scheme. The proposal commonly induces a discussion among mappers before commencing a vote on whether or not to endorse the tag. We analyse the proposals and the community discussions for each endorsed tag in our research database to infer motives for endorsements by the platform. Via the history function implemented in the wiki environment we could track changes in past versions of

³¹ Also see the appendix

proposal pages and tag presentation. This allows us to record the exact point in time when each tag was first endorsed and get an understanding of the initial rationale among stakeholders that lead to the endorsement of a tag.

We conduct a content analysis of the tag endorsement proposals aimed at the structured identification of patterns (Schreier, 2014). The analysis broadly followed four steps. First, we systematically documented all endorsed tags in our sample (n = 270). Second, we examined the tag proposal pages on the community wiki (described below) and noted down keywords and excerpts from the tag proposal discussions that pointed to why the tag was considered for endorsement. This step resulted in a list of some 15 generic themes describing motivation for each tag's endorsement. Salient themes were, for instance, 'establish new tag category', 'delineate alternatives', 'increase detail of the map,' or 'increase global applicability of tag.' Third, in a subsequent step we cross-referenced these emergent themes with observations made in studies on platforms in top-tier journals. As such, we paid close attention to notions of strategies, competitive moves, or advancement of innovation on digital platforms. The goal of this step was to corroborate the formulated themes and achieve convergence with categories that are meaningful to platform scholars. Fourth, informed by insights in the literature, we returned to the content analysis and aggregated the generic themes into strategic motives of endorsements. We surmise that such endorsements are undertaken to support strategies known in the platform literature. In the following subsection, we highlight how the resulting motives align with strategic actions in the platform literature.

4.4.2. Endorsement Motives

During the time frame covered by our study, 270 tags from our sample were endorsed on the platform. All tag endorsements share the intention of increasing the level of detail the platform can present and giving direction by delineating alternative metadata descriptions. However, beyond this commonality, the rationales as reflected in the tag proposals reveals differences as to why certain tags were endorsed. Specifically, we identified four motives of endorsements. In this section we present the identified motives and demonstrate that these resonate with strategies by platform operators put forth in the literature yet remain empirically underdeveloped. Linking prior theoretical contributions with the motives of endorsements identified here will help us create

insights into the effect of endorsements to support established strategies in the platform literature.

Endorsement Motive 1 - Commit to New Market

Evidence from our exploratory analysis of the qualitative empirical material reveals a clear pattern in the motives of tag endorsements on OSM. Specifically, two tag groups (`building=`, `shop=`) are examples of endorsing a new category in order to fully commit the platform to a new market segment. While a first `building` tag existed since 2007 and the endorsement of an initial generic category (`building=yes`) has been discussed since 2008, OSM mappers were reluctant to add building descriptions in fear of diluting the focus of a general purpose map. However, with rising competition, the platform engages in what we refer to as “commitment to a new market” and endorses building tags with more details to open up the category for use on the map. For instance, in 2014 during which most building tags were endorsed, one community member summarises the decision as:

“If we do not enter more details into the map (e.g. because we use too generic [tag] types) we will limit the possible use cases and reduce the information in the database.” -- User on OSM Wiki, March 2014

Tags of `shop=` category followed a very similar pattern. Metadata descriptions detailing commercial operations were long disputed on the platform. However, with a growing level of detail on the map, and an increasing pressure to ensure the map remains useful, the decision was taken to include a new `shop=` category. As one user remarks on including shops in addition to the existing amenity tags:

“I support a new `shop=<shop category>` tag. [...] It makes things easier than sorting through `amenity=` tags for shops” -- User on OSM Wiki, October 2006

Similarly, another user states:

“I think the proposal of separating this tag from the amenity stuff is a good idea. I think it would be better to have a separate category for shops. `shop=bakery`, `shop=butcher` etc. Then shops can be rendered with a generic icon and the system doesn't have to know about every kind of shop.” -- User on OSM Wiki, December 2006

Defining a platform’s market footprint, that is, which market the platform addresses and what it offers, is a crucial decision for platform operators. The platform literature has thereto proposed several pathways including opening the platform and granting

access to a new party (Boudreau, 2010), or expanding the platform by offering new complementary products (Eisenmann *et al.*, 2011). Boudreau 2010 refers to allowing a new category of complementors onto the platform as ‘opening up markets’ -- the consequences of which are an increase in the number and diversity of products offered on the platform (Boudreau, 2010). Similarly, Eisenmann *et al.* (2011) demonstrate how expanding the product categories offered on platforms is a powerful competitive move to ‘envelop’ and occupy complementary market segments or compete in rivals’ purviews.

In the literature on multi-sided marketplaces endorsements that signal future activity are associated with network effects (e.g., Dranove and Gandal, 2003). With the intention to ‘open up markets’ for the platform (Boudreau, 2010), endorsing a commitment to a new market is a strong signal to platform participants. An endorsement with this motive would make the direction of the platform explicit by signalling what kind of interaction is desired on the platform.

We identified two cases of endorsements signalling the commitment to a new market, 44 endorsements of tags in the building category and 80 endorsements of tags in the shop category. We hypothesise that endorsements that signal ‘a commitment to a new market’ have positive effects on both aspects of platform generativity:

H1a: An endorsement by the platform operator that signals to commit the platform to a new market increases platform scale.

H1b: An endorsement by the platform operator that signals to commit the platform to a new market increases platform scope.

Endorsement Motive 2 - Accommodate Third-Party

OSM geo-data is used by a variety of external applications. These range from specific mapping services (e.g., routing and navigation), to more general purpose applications using geo-spatial information as part of their product (e.g., Foursquare). External application providers take an interest in the OSM tagging scheme, and especially in the endorsements by the platform. Having the platform endorse a tag that is needed

for the service of a specialised application, ensures continued smooth operation of that third-party offering.

Through our in-depth exploration of tag endorsement proposals, we have identified 23 endorsements that accommodated third party requests. These endorsements were either directly motivated by an external application and its usage of a specific tag, or members of the community made strong cases that endorsing a certain tag would lead to additional external use cases.

In one example, the proposal page for endorsing the tag *highway=priority* states:

“This [tag] will help a safety app or automated vehicle know that emerging side-road vehicles may create a hazard at that main-road junction.” -- Presentation of the tag in the OSM Wiki

Other users claim benefits for external apps, for instance, the tag *highway=traffic_calming*:

“Having just been routed by my SatNav down a road with half a dozen speed bumps even though it is marked on maps as a tertiary route, I was struck by the need to have this information in the map data even if it is not displayed, to that these ways can be avoided” -- User on OSM Wiki, December 2006

In another case a platform participant from a third party app argues:

“A reasonably new feature for OpenTripPlanner [an external application] is the ability to read highway=elevator nodes. When these nodes are connected to ways with levels [...], they should be interpreted correctly.” -- OSM Wiki page of the external application, April 2012

The dynamic of endorsements such as the above echoes a prevalent debate in the literature on platform strategy. Central to which is the challenge of reconciling the interests of platform operator with the requests by its participants (Wareham *et al.*, 2014, Parker *et al.*, 2017). It is in the platform’s interest to provide guidance as to what forms of interactions are desired on the platform. Conversely, external participants seek flexibility to create derivative products and services according to their own strategy. Eaton *et al.* (2015), for instance, have pointed out how the design of platform components develop beyond initial intentions by the platform operator as platform participants seek to have their interests reflected in the platform’s functionality (Eaton *et al.*, 2015). They describe this evolution as a tuning process where components on the platform boundary are shaped by "cascading actions of accommodations and rejections of a network of heterogeneous actors and artifacts" (ibid:p. 217). Reconciling these divergent interests is a challenge for smoothly operating a platform.

Pertinent solutions to this have often involved arrangements to govern the relationship between operator and external participants. West (2003), for example, relates the episode of IBM adopting Linux. While reluctant at first, IBM later decides to endorse the use of Linux software on IBM servers in reaction to a ‘market pull’ by internet service providers (ibid; p.1273). Accommodating for the use of open source software led to novel kinds of platform participation from external actors (West, 2003).

Endorsing tags so as to accommodate third party requests is akin to predictable standards and as such “provide complementors with assurance of the eventual scalability and reuse of their innovations” (Wareham *et al.*, 2014; p. 1200). We thus hypothesise that endorsements that signal the accommodation of third party requests benefit platform generativity:

H2a: An endorsement by the platform operator that signals to accommodate a third party increases platform scale.

H2b: An endorsement by the platform operator that signals to accommodate a third party increases platform scope.

Endorsement Motive 3 - Balance Market Demand

In the context of OSM, the notion of ‘tag anything you see and use any tag you like’ is paradigmatic (Ramm *et al.*, 2010). The level of detail the map provides is determined by what mappers, users, and developers wish to have represented in the geo-data. This often leads to situations in which a number of similar tags are used to describe the same category of object in the real world. Typical examples of this are tags that use very local nomenclature (e.g., highway=motorway, or amenity=biergarten) or tags that present rival descriptions for essentially the same object (e.g., shop=ice_cream, vs. amenity=ice_cream). This lack of consistency is a problem for the platform as it makes the use of data objects seem arbitrary and subject to local circumstances. Eventually, this may prevent usage of the database from growing. Here, the platform often endorses tag descriptions to signal that the future direction ensures a balanced yet sufficient demand in the platform.

During our content analysis of tag endorsement proposals, we have identified 27 cases where the endorsement signalled to create balance among the audience of the platform regulating alternative options. Indicators for endorsements of this motive often

included comparing usage number of rival alternatives, or seeking to endorse the tag with the greater global applicability.

For example, one endorsement proposal draws from a database query and claims:

“On the basis of this evidence [a database query] it should be documented that amenity=doctors is [already] the most commonly-used way to tag a doctor's office in practice. [and hence this option should be endorsed]” -- User on OSM Wiki, November 2010

One user remarks clearly:

“Some tag-proposals for specific facilities exist, but the more general ‘social_facility’ tags will allow easier organisation and searching from a single top-level feature.” -- Description of the tag proposal on OSM wiki, July 2010

The discussion on highway classifications exemplifies the difficulty of agreeing on tags that are globally applicable while remaining locally accurate:

I'm not convinced that most [highway=]trunk roads are dualled. In the UK we tag the "primary route network" as trunk. The majority of that are single carriageway. Oneway is therefore not implied, and must be added as a separate tag. -- User on OSM Wiki, June 2008

Platform operators need to balance demand in the platform in a such a way that platform usage remains general enough to attract new participants, while retaining existing ones. For instance, if participants contribute applications to a software platform, it is imperative for platforms to continuously enrol new external contributions in order to remain successful. Parker *et al.* (2017) analyse the trade-off decisions that platform operators face when attracting developers. All things being equal, platforms strive for larger audiences of participants than smaller ones as platform competition increases (Parker *et al.*, 2017). Maximising the installed base of a platform is itself a potent signal to platform participants as it assures them of sufficient size of the platform for their services (Boudreau and Jeppesen, 2015; Brynjolfsson and Kemerer, 1996). However, the number of product alternatives offered on a platform, also need to be carefully balanced. Boudreau (2012), evidences an effect of what he refers to as ‘overcrowding’ - describing the innovative activity on the platform as a function of the number and diversity of application that exist simultaneously. Avoiding that the sheer number of possible interactions on a platform

induces negative effects echoes the notion of managing desirable and undesirable outcomes of generativity (Wareham *et al.*, 2014). We thus hypothesise that endorsements aimed at signalling a balance in demand have positive impact for platform generativity.

H3a: An endorsement by the platform operator that signals a balance in market demand increases platform scale.

H3b: An endorsement by the platform operator that signals a balance in market demand increases platform scope.

Endorsement Motive 4 - Ratify Emergent Use

From our analysis of tag proposals, we have identified 23 cases in which the endorsement was a reaction to use cases already established on the platform. This pattern is distinct from the ones mentioned above in several ways. It differs from a third party request as this pattern is not motivated by external applications. It is also distinct from balancing market demand as it is not aimed at ensuring sufficient audience sizes. Instead, endorsements with this motive acknowledge that a use case has developed on the platform that now needs to be supported in order to retain momentum. The cases we identified most typically saw mappers sanctioning emergent use cases for example by recognising available tag combinations that were not initially anticipated.

On *railway=subway_entrance* a mapper describes the motivation for the endorsement proposal:

“I created this page as this tag is currently in use but poorly documented.” -- User on OSM Wiki, July 2012

Similarly on the tag *amenity=motorcycle_parking*;

“This is a pretty well used tag nowadays. Of course amenity=motorcycle_parking was a natural thing to add following after the amenity=bicycle_parking tag” -- User on OSM Wiki, April 2016

Platform operators often face trade-offs between the use cases that were anticipated, and the use cases that are eventually enabled on the platform. Operators are therefore often compelled to react to newly generated interactions resulting from actors engaging with each other and the platform. Reacting to such emergent use cases is often motivated by a desire to fuel further activity on the platform. An example of this is Google’s decision to offer its own Photo App after platform participants

demonstrated sufficient demand in such services (Förderer *et al.*, 2018). In their study, Förderer *et al.* (2018) show how this decision resulted in continued innovation on the platform as Google implicitly endorses an emergent application domain. Another instance is reported by Wareham *et al.* (2014). They describe how predefined implementation methodologies by a large enterprise software vendor are amended by partners to facilitate specialised instantiations of platform products in local markets. The vendor acknowledges that peripheral participants deviate from suggested practices and supports the idiosyncratic alteration of its procedures *ex post* (Wareham *et al.*, 2014). We thus hypothesise:

H4a: An endorsement by the platform operator that signals a ratification of emergent use increases platform scale.

H4b: An endorsement by the platform operator that signals a ratification of emergent use increases platform scope.

Table 4.1 below summarises the motives for endorsements we have identified in the data and their connection to decisions by platform operators in the literature. We embrace the fact that the identified motives are reflected in literature on platform strategy as it demonstrates the applicability of these endorsements for platform scholarship. In the same time, it underlines our motivation of studying endorsement signals as potential triggers of generativity. While the motives in platform strategy are clear, the effects of endorsements are not.

Table 4.1. Identified Motives for Endorsement			
Motive	Description	Example from Content Analysis	References
Commit to New Market	Expanding the market footprint of a platform by establishing a categorically new way of interacting with platform resources.	Endorsing the <i>building=</i> and <i>shop=</i> tag categories as new additions to the data model	Boudreau 2010 Eisenmann <i>et al.</i> 2011 Evans <i>et al.</i> 2009

Accommodate Third-Party	Accommodating interests of platform participants	Endorsing <i>highway</i> =tags in the database that provide finely grained details to external navigation applications	Eaton <i>et al.</i> 2015 West 2003
Balance Market Demand	Directing the platform such that the installed base is maximised	Endorsements that prioritise general and widely applicable tags over niche or localised alternatives	Parker <i>et al.</i> 2017 Boudreau 2012
Ratify Emergent Use	Reacting to unanticipated use cases	Endorsing tags that were unanticipated such as detailed distinctions for rental objects (boats, cars, bikes, etc.)	Förderer <i>et al.</i> 2018 Wareham <i>et al.</i> 2014

In summary, 197 of the 270 tag endorsements correspond to the motives we suggest above. The remaining tags are not distinctly attributable to one of the motives. Either because the tag proposal provided no information at all, or the information content was so limited that a motive was not discernible. However, these tag endorsements are valuable for our study. They share the objective of increasing the level of detail of displayed map content, as well as providing clarity and direction for the platform. Endorsements that do not present a dedicated motive can therefore act as a reference group for the efficacy of increasing scale and scope through the four identified endorsement motives and we include these endorsements later in the analysis.

4.5. ANALYSIS STAGE 2: CONFIRMATORY HYPOTHESIS TESTS

4.5.1. Data

The study draws from a full OSM database excerpt in the time from May 2009 to December 2014. The time frame was chosen to reflect data model changes in April 2009 hence ensuring consistency in the underlying data structure (Ramm *et al.*, 2010).

For the analysis we obtained a full history extract of all database objects ever created in Europe until January 2015. To focus our analysis, we selected ten high level, general purpose tags which are foundational to mapping geographic features, yet are also useful for external applications using OSM data. We provide a full overview in the appendix.

We use the database excerpt to generate two data sets, each focusing on one aspect of generative change; scale and scope. To test the effects on scale, we use information on the usage of each tag. Specifically, we count the number of database objects associated with each tag per month between May 2009 to December 2014. To test the effect on scope, we randomly sampled tagged OSM objects over the same period. This information is used to compute the number of tags on each object and thus reflects the level of detail of the map. The data in both samples run for a period of 66 months.

We next analyse how each endorsement motive affected scale and scope of the platform. We regard an endorsement signal as successful in triggering generative platform change if it increases both scale and scope of the platform. Our unit of analysis is thus an individual *key=value* pair in the database that we simply refer to as tag. We are interested in two metrics for any given tag. First, the number of objects that the tag describes. Second, the level of detail that an endorsed tag provides.

Dependent Variables

We use two dependent variables to capture scale and scope of the platform respectively.

1. *Tag Usage Growth: To assess platform scale we use the growth of tag usage measured in the monthly percentage increase of the number of objects each tag (key=value combination) describes in the database.*

2. *Tag Occurrence: To assess platform scope we measure the number of metadata descriptions on data objects. Such a measure is commonly referred to as occurrence and captures the number of tags that appear in addition to the focal tag.*

Controls

We include a number of control variables to rule out alternative explanations. In particular, we control for the size of the database, the time passed since a tag was endorsed, and the amount of tags in one group that is endorsed already. We additionally include time and tag fixed effects. Table 4.2 below gives an overview over all constructs and variables used in this study.

Table 4.2. Variable Definitions	
Variable	Definition
1a Tag Usage Growth	The monthly percentage increase of the number of objects that each tag describes.
1b Tag Usage (binary)	A binary transformation of variable 1a that is 1 if the mean growth rate is above the platform average of 4.4%, and 0 otherwise
2 Tag Occurrence	The number of tags occurring on a data object together with one of the focal tags
3 Objects in Database	The total number of geo-data points in the database excerpt of Europe
4 Tag Tenure	The number of months since a tag has been introduced and discussed by the core community
5 Tags Documented	The percentage of tags per group (amenity=, building=, etc.) that are currently documented for a proposal process i.e. are discussed on the platform
6 Tag Fixed Effects	Dummy variables denoting the top level tag group for each observation (amenity=, highway=, etc.)
7 Time Fixed Effects	Dummy variables denoting the month/year of observations

For the quantitative analysis of scale and scope we implement a modelling approach that offers high flexibility and interpretability while being thoroughly grounded in methodological convention. To analyse the effects on platform scale through tag endorsements (H1a through H4a), we use a logistic regression to assess if endorsed tags are likely to grow faster than the platform would on average. For our investigation of the effects on platform scope through tag endorsements (H1b through H4b), we use a quasi-Poisson model to compare the effect of an endorsement on the level of detail each tag provides before and after it is endorsed. In the following section we detail our modelling approach for each aspect of generative change before presenting the results.

In both cases we draw from the generalised linear modelling framework to enable the analysis of non-normal data structures through linear inferences while increasing interpretability (Hardin and Hilbe, 2003; Hosmer, Lemeshow, and Sturdivant, 2013).

4.5.2. Endorsement and Platform Scale

To assess the effect of platform endorsements on increases in platform scale, we estimate the likelihood that an endorsed tag will grow faster than the platform average. We therefore computed the monthly growth rate, that is, the percentage increase in the number of objects each tag described between May 2009 and December 2014. To account for the underlying trend of the platform we centred the monthly growth rate of each tag around the average growth of the entire Europe database in the same time window (4.4% per month).

Consistent with the logistic regression framework, we then transformed the average tag usage growth rate into a binary variable that is 1 if the mean of the monthly usage growth rates of an endorsed tag is positive (i.e. on average higher than the platform trend), and 0 if otherwise.

We first compare how often tags of each motive for endorsement exhibited above average growth before and after an endorsement. To make this comparison meaningful, we use the performance of unendorsed tags as a baseline contrast. As such, one in five tags (18%) grows faster than the platform average without ever being endorsed. With this baseline we use a chi-square test of independence to examine the effect on platform scale in each endorsement motive (i.e. the number of tags with above average growth). The relation between endorsement and above average growth is significant endorsements with motives ‘Commit to New Market, ($X^2 = 99.527$ $df = 1$, $p < 0.001$), ‘Accommodate Third Party’ ($X^2 = 10.115$, $df = 1$, $p < 0.001$), and marginally significant for ‘Balance Market Demand’ ($X^2 = 2.4741$, $df = 1$, $p = 0.057$). These motives were thus more likely to include tags whose usage outgrew the platform average after they had been endorsed. The test however, was not significant for endorsements with the motive to ‘Ratify Emergent Use’ ($X^2 = 1.349$, $df = 1$, $p = 0.8773$).

We next model the relationship between endorsements and above average growth as a logistic regression and contrast the effects of endorsed tags from each of the identified motives to the baseline group of unendorsed tags. To ensure independence among observations, the tag groups have been crisply defined and are thus perfectly mutually-exclusive.

Despite the stratification in contrast groups, the sample size remains sufficient for log odds estimation of the effect after the endorsement. This is due to a non-rare target event (tag growth > platform average; $\text{PrE}_{\text{min}} = 8.7\%$ *Ratify Emergent Use*, $\text{PrE}_{\text{total}} = 39\%$) and the low multicollinearity among selected variables in the model ($r_{\text{max}} = 0.203$) (Bergtold, Yeager, and Featherstone, 2018). Other than that, logistic regression models make few assumptions about predictor properties. This provides a flexible and interpretable way to compare the effect of categorical predictors (Hosmer *et al.*, 2013). We applied AIC stepwise forward variable selection (Grogan and Elashoff, 2017) and likelihood ratio tests to confirm goodness-of-fit (see e.g., Hosmer *et al.*, 2013) for the full model ($X^2 = 91.7$, $df = 23$, $p < 0.001$). We model the probability of a tag endorsement to outgrow the platform average as:

Y(logit) Tag Usage Growth > Platform Average = $\beta 1$ motive for endorsement + controls.

4.5.3. Endorsement and Platform Scope

To test if endorsements also increase the potential scope of the platform, we analyse the number of metadata descriptions that occur together with tags in each pattern before and after a tag is endorsed.

Our variable of interest consists of count responses. These counts represent the recorded tag occurrence, that is, the number of metadata descriptions that appear on each data object. Rooted in the empirical context, tag counts are discrete and non-negative; while OSM objects with no tags are very rare in the entire OSM database, they were impossible in our sample since objects are included on the basis of having at least one tag (e.g. *highway=*). This results in a non-normal distribution of all tag counts in the sample. In the absence of normality and the inadvisability of transforming count data, the analysis is conducted using Quasi-Poisson regression – a method that is widely popular in ecology (e.g., Bolker, 2008) and well documented in

the statistics literature (Hilbe, 2011, 2014; Winkelmann, 2008). Specifically, we model the expected response in the group means (Hardin and Hilbe, 2003; p. 56). As often the case with real life count data, our sample does not adhere to *equipersion*, that is, the variance does not equal the mean and overdispersion is consistently greater than 1 (Hilbe, 2011).

Initial exploratory analysis of our sample supports the applicability of a Quasi-Poisson approach. For instance, the linear relationship between sample mean and variance indicates quasi-model specifications as the preferable choice. Under linear mean-variance conditions in overdispersed count data, Quasi-Poisson ensures more accurate estimates compared to possible underestimation of the variance when assuming normality, or overestimation when assuming quadratic relationships inherent to negative binomial distributions (Hilbe, 2011; Ver Hoef and Boveng, 2007). Indeed, quasi-likelihood tests (Bolker, 2016; Ver Hoef and Boveng, 2007), of our base models yield a >30% increase in precision for the Quasi-Poisson model compared to stricter Poisson models, indicating that a flexible treatment of dispersion is needed to describe the data (Hilbe, 2014).

Quasi-Poisson models have been demonstrated as efficient and flexible alternatives to more complex modelling approaches (Ballinger, 2004) and the approach is particularly valuable for our research design for two reasons.

First, Quasi-Poisson models extends the Poisson distribution through dispersion parameter estimates that account for the extra variation relative to Poisson (Hilbe, 2011, 2014). Second, as a non-parametric technique, the approach does not require a probability distribution function. Instead, the covariate effect on mean variations is analysed. This is adequate in our research design as we are interested in the differences of the average number of metadata for each endorsement motive. In summary, using Quasi-Poisson offers us a robust and interpretable method whose estimates account for overdispersed count responses with high flexibility (Hilbe, 2014). With $Y_{count} \sim Poisson(a)$, the model is thus denoted as:

$$Y_i \text{ Tag Occurrence} = \alpha\mu + \beta_1 \text{ motive for endorsement} \times \text{time of endorsement} + \text{controls}$$

The model analyses the average number of tags associated with an OSM object. Where Y is the average number of tags on objects with tag i , μ the estimated intercept given the dispersion parameter α , and β_i the coefficient for the contrasts between the identified motives before and after their endorsements. For a baseline reference group, we sampled OSM objects with attributes that are representative of the OSM database.

4.6. RESULTS

We report the results of our hypothesis tests in pairwise sequence. Tables 4.3 and 4.4 detail the respective regression results from the full models.

Hypothesis 1 states that endorsements aimed at increasing a platform's market footprint increases both scale and scope of a platform. Informed by the platform literature, we refer to such moves as a 'Commit to New Market' endorsement that signals that a platform addresses a new market segment or offers a new product category. In the test of hypothesis 1a, the coefficient in the logistic regression is positive ($\beta = 2.380$) and highly significant ($p < 0.001$). This implies that tags endorsements of this form are ~10.8 times more likely to grow faster than the platform average compared to the reference group of tags that are never endorsed. This confirms hypothesis H1a. For the analysis of platform scope through 'market commitment' endorsements, we find that the level of detail on objects with such tags not only increases after the endorsements, but that the average number of metadata is higher than the platform average (coefficient after endorsement: $\beta = 0.197$, $p < 0.001$). This confirms hypothesis H1b.

Hypothesis 2 claims that endorsements that signal the accommodation of third party requests increase both scale and scope of the platform. The literature suspects such actions to be necessary and beneficial to generativity on platforms (eg. Eaton *et al.* 2015). Our analysis aligns with that suspicion as both hypothesis 2a and 2b are confirmed. As for scale, (H2a) tag endorsements as a reaction to third party requests show the starkest contrast compared to the baseline group of unendorsed tags; the coefficient is positive and significant ($\beta = 2.513$, $p < 0.01$). With regards to gaining scope (H2b), endorsements in the form of 'Accommodate Third Party' perform similarly well. The coefficient of the average tag occurrence after endorsement is

positive and highly significant ($\beta = 0.281, p < 0.001$) indicating the strongest increase in scope of all tag endorsements.

In hypothesis 3 we surmised that endorsements signalling a balance in demand across the platform have a positive effect on both scale and scope. While we could confirm the effect on scale (H3a) of such endorsements (the coefficient is positive $\beta=1.486$ and significant at $p < 0.05$), our analysis shows a decrease in the level of detail after tags in this category have been endorsed ($\beta = -0.159, p < 0.001$). On that basis, we reject hypothesis H3b as the level of detail on data objects with tags endorsed so as to ‘balance market demand’ does not increase with endorsement.

Lastly, in hypothesis 4 we assumed a positive effect on platform scale and scope of endorsements as ratifications of emergent use cases. We were unable to confirm either assumption as both effects are not significantly different from the platform average (scale; $\beta = 0.471, p > 0.1$; scope: $\beta = 0.013, p > 0.1$). This indicates that the effects of such endorsements did not significantly differ from average activity on the platform in our sample. We therefore reject hypotheses 4a and 4b.

In summary, our results indicate that endorsed tags are more likely to grow in scale faster than the platform would on average (H1a-H3a). Exception to this are endorsements with the motive to ‘Ratify Emergent Use’ (H4a), as tags are not significantly different from tags that are never endorsed. Figure 4.3 below illustrates the results from the logistic regression. Clearly visible is the difference in the likelihood of endorsed tags outgrowing the platform average in comparison to tags that are never endorsed.

Figure 4.3. Illustrated Results from Logistic Regression

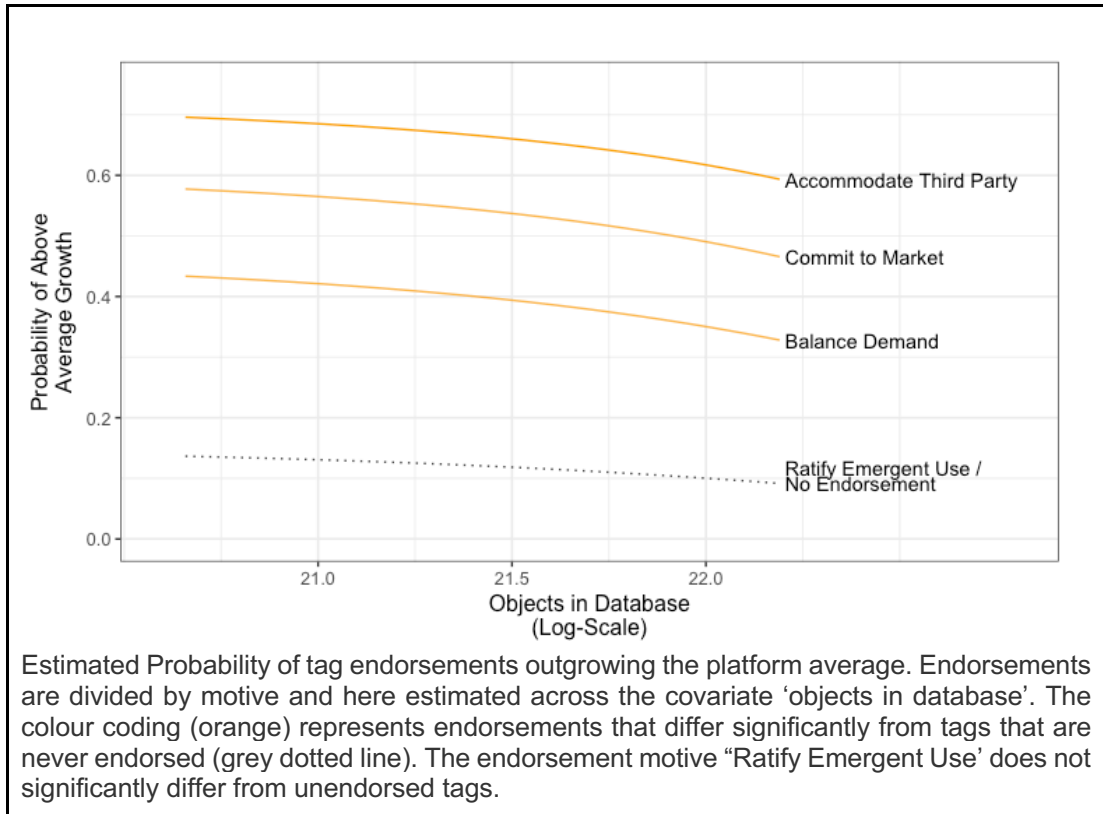


Table 4.3. Model Results – Platform Scale (Logistic Regression)		
Dependent variable:		
Tag Growth > Platform Average Binary		
Motive: Commit to New Market	2.380***	(0.617)
Motive: Accommodate Third Party	2.513**	(0.862)
Motive: Balance Demand	1.486*	(0.626)
Motive: Ratify Emergent Use	0.471	(0.882)
Objects in Database	-1.312e-9*	(7.087e-10)
Tag Tenure	-0.064***	(0.012)
Tags Documented (% in Group)	-11.32	(18.44)
Contrast (Other Endorsed Tags)	included	
Tag Fixed Effects	included	
Time Fixed Effects	included	
Intercept	0.053	(0.948)
Observations		370
McFadden Pseudo R-Squared (%)		26.5
Counts Pseudo R-Squared (% at .50 cutoff)		0.77
AIC/AIC (controls only model)		409.4 / 451.6
Note:		
*p<0.05; **p<0.01; ***p<0.001		
Robust standard errors in parentheses; coefficients in log-form		
Pseudo R Squared Measures are based on 10 cross-validations		
using samples of 100 objects with 'unendorsed' tags for baseline reference		

The results of our tests of H1b and H2b indicate that endorsements signals with motives 'Commit to New Market' and 'Accommodate Third Party' increase

significantly in the level of detail they provide to the platform and hence contribute to the scope of the platform.

In contrast, tags endorsed so as to ‘Balance Market Demand’ across the platform decrease in the level of detail in our sample. These tags thus fail to increase platform scope.

With regards to H4b the picture is mixed. While the tag improves in the level of detail, it remains around the platform average, and tags endorsements that ratify emergent use cases are not significantly different from the baseline group of the platform average, thus not contributing substantially to the scope of the platform. Figure X below visualises the quasi-Poisson results before and after tags have been endorsed. The dotted line represents the average level of detail in the metadata across the platform.

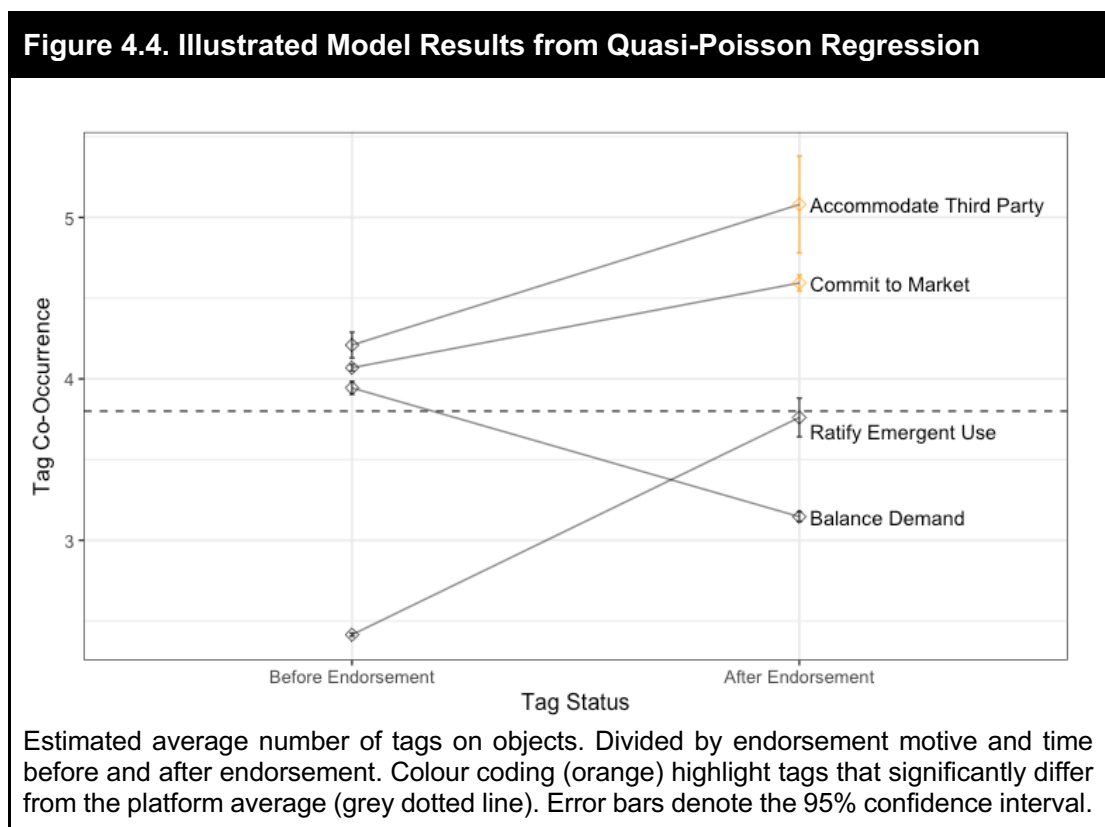


Table 4.4. Model Results – Platform Scope (Quasi-Poisson Regression)		
	<i>Dependent variable:</i>	
	Tag Occurrence	
	<i>Discrete, Non-Negative</i>	
Motive: Commit to New Market (after endorsement)	0.197***	(0.014)
Motive: Commit to New Market (before endorsement)	0.066***	(0.015)
Motive: Accommodate Third Party (after endorsement)	0.281***	(0.025)
Motive: Accommodate Third Party (before endorsement)	0.086***	(0.016)
Motive: Balance Demand (after endorsement)	-0.159***	(0.011)
Motive: Balance Demand (before endorsement)	0.059***	(0.011)
Motive: Ratify Emergent Use (after endorsement)	-0.013	(0.019)
Motive: Ratify Emergent Use (before endorsement)	-0.460***	(0.013)
Objects in Database	7.396e-11***	(1.115e-11)
Tag Tenure	-0.008***	(0.002)
Tags Documented (% in Group) ¹		included
Contrast (Other Endorsed Tags)		included
Tag Fixed Effects		included
Time Fixed Effects		included
Intercept	0.574***	(0.048)
<i>Observations</i>		97,890
<i>McFadden Pseudo R-Squared (%)</i>		10.2
<i>quasi AIC / AIC (controls only model)</i>		91,175 / 98,123
<i>Note:</i>		
*p<0.05; **p<0.01; ***p<0.001		
<i>Robust standard errors in parentheses; coefficients in log form</i>		
¹ variable included as quartile split to avoid multicollinearity		

We regard an endorsement as a successful signal if it increases both scale and scope of the platform. Table 4.5 below gives an overview over our hypothesis test results.

Table 4.5. Overview of Hypothesis Test Results		
Endorsement Motive	Aspects of Generative Change	
	Increases Platform Scale	Increases Platform Scope
Commit to New Market	confirmed (H1a)	confirmed (H1b)
Accommodate Third Party	confirmed (H2a)	confirmed (H2b)
Balance Market Demand	confirmed (H3a)	not confirmed (H3b)
Ratify Emergent Use	not confirmed (H4a)	not confirmed (H4b)

4.6.1. Robustness Checks

We applied a battery of robustness checks to ensure confidence in our surmised effects. Initially, both models were controlled for three major confounders.

First, the 2012 change-over to an Open Database License affected the logic of how and which data objects on OSM are available for re-use. The change commenced in July 2010 after which every user was automatically signed up to a new license agreement. From September 2012 onwards the new license was in place and objects that were not certified for re-use were automatically redacted at this point. As a major external shock to the platform, the license change influenced all data use and had potential implications for both predictors and outcomes in our used models. We therefore re-ran the analyses in both models using the above dates as cut-offs for subset analyses of objects affected by different licenses. Using the full models, we achieved consistent results for all effects, assuring us that the effects reported here are not affected by the license change.

Second, using the OSM wiki as the main source for our classification of tags, requires to control for possible confounding due to *de-facto* endorsements by the community. While tags are typically endorsed by majority vote, some tags are endorsed in less transparent de-facto decisions. That is, community members unilaterally decide to change the status of a tag such that it appears as if it has been collectively endorsed, although no majority vote was held. We thus cross-referenced tags that are listed on the community's 'approved map features' wiki page to control for measurement errors arising from statuses that may well be endorsed as de-facto standards but did not undergo voting processes. We re-ran the analyses on these subsets using the full models. We achieved consistent results for the logistic regression model. For the Quasi-Poisson model, the effects for the endorsement motives, 'Commit to Market', and 'Third Party Request' are consistent. For tags in the group 'Balance Demand' and 'Ratify Emergent Use' the coefficients remain unaffected by the endorsement albeit being marginally positive. Despite this deviation from the full model on the full data set this does not contradict our surmised effects about the efficacy of the different endorsements.

Third, we ran subset analysis on the tenure of tags, that is the time each tag was documented and discussed by the community. A source for potential bias here is that

tag endorsement impacts for scale and scope are merely functions of the age of a tag. For instance, objects with seasoned tags could attract more additional tags, or vice versa, objects with newer tags could be richer in description assuming that the tag fills an awaited need for stakeholders. We re-ran the analysis with different cut-offs of the tenure of tags and achieved consistent results.

To address concerns of endogeneity, we implemented two post-hoc tests to rule out that the explanatory treatment variable ‘endorsement’ is endogenous. We therefore applied a control function procedure following Rivers and Voun (1988) and Train (2009) as well as an instrumental variable approach (Wooldridge, 2002). Results of both tests are inconspicuous of endogeneity (see details in the appendix).

Lastly, we ran both models with standardised coefficients and achieved consistent results, we included alternative control variables (such as number of key=value pairs, or number of users over time) and achieved consistent results.

4.7. DISCUSSION

Generativity is a highly desirable goal for platform operators as it sustains growth and innovation on digital platforms (Yoo *et al.*, 2010, Wareham *et al.*, 2014; Boudreau, 2012; Parker *et al.*, 2017). Yet, there is little clarity about how generativity can be stimulated in areas where it is desired by the platform operator. Since platform participants have no way of knowing where generativity is desired, a fundamental challenge is the information asymmetry between the platform and its participants (cf. Ho and Rai, 2017; Boudreau and Jeppesen, 2015; Evans and Schmalensee, 2016). Signalling the platform operator’s intentions might be a significant way for platforms to stimulate generativity in areas of interest. To this end, we set out to study how and why signals in the form of endorsements by the platform operator can stimulate generative change. Zooming in on specific rationales of endorsements, we derived and tested four motives of endorsements and their impact on generativity in the context of the geo-data platform OpenStreetMap.

Contrasting the four different motives of endorsements, our findings give insights into how, and why signals by the platform operator can stimulate generativity. In essence, signalling addresses information asymmetries between senders and receivers of

information. In the context of digital platforms this is a potentially powerful lever. Signals make intentions of the platform operator known to platform participants (cf. McIntyre and Srinivasan, 2017). Signalling also instils confidence in otherwise unobservable qualities of a platform - such as potential for future growth (Boudreau and Jeppesen, 2015) or predictability for participant contributions (Ho and Rai, 2017).

In order for platform operators to successfully signal which interaction is desired and where, two aspects are crucial for the efficacy of a signal. First, the underlying quality that is intended to be signalled needs to be correctly identified. Since platform participants cannot know for certain which kind of interactions on the platform are desirable, the platform operator has to identify this quality and make it known. Second, a signal needs to be easily understood by the receiver. That is, the effort required to decode the signal is best limited so as to facilitate information acquisition on the receiver's end (Connelly *et al.*, 2011). Answers as to why certain endorsements work while others do not, can therefore be framed as a function of friction in the exchange of information between platform operator and platform participants.

4.7.1. Successful signals of endorsement

Our first set of findings relates to the two successful forms of endorsements in our analysis. We find that endorsements by the platform operator are supportive of generativity if the signalled quality has been correctly identified by the platform operator and is communicated in a way that is easily observable for platform participants. In other words, the endorsement signals an aspect of platform interaction that is desirable for both platform operator and platform participants.

In the case of endorsements that signal a commitment to a new market, we find clear evidence that the interaction is highly desirable for the platform. The content analysis revealed that a clear motivation exists to enable more and novel interactions by committing to a new tag category so as to sustain growth and innovation of the platform. Once endorsed, the presence of the platform in a new segment of the market is clearly recognisable for platform participants. By signalling an extension of the possibilities that exist to engage with the platform, the platform operator explicates design choices (Baldwin and Clark, 1997; Tiwana, 2014). The authority that is indicated by an endorsement (Budhathoki and Haythornthwaite, 2012) implies that

desirable variation has been identified as the underlying quality which is now being signalled to participants (Wareham *et al.*, 2014). This is akin to a competitive move of the platform into a complementary market segment (Eisenmann *et al.*, 2011). As such, the endorsement signals a future growth in the installed base. This expectation about the future state of a platform creates network effects (e.g., Fuentelsaz, Garrido, and Maicas, 2015).

In the case of accommodating third party requests, the endorsement signals that the platform embraces interaction with external platform participants. Explicating this otherwise barely observable quality is crucial for a signal and directly speaks to the “unobservable ability of the signaller to fulfil the needs or demands of an outsider observing the signal” (Connelly *et al.*, 2011, p. 43).

This endorsement motive presents the best performance in our sample. This is noteworthy as it demonstrates that the ability and willingness of a platform to accommodate requests from third party participants can be confirmed by the right signal from its operator. A signal in turn also clarifies what the platform operator envisages interaction with the platform to look like and with whom. This demonstration alone can be seen as a “signalling device” (Evans and Schmalensee, 2016) in that an endorsement with this motive signals that third party engagement is desired in the first place and requests by participants are being accommodated. This also signals predictability in the platform’s approach to third parties and the platform operator thereby instils confidence in participants that their services can draw on platform resources in the long run (Wareham *et al.*, 2014).

In that sense, the first set of findings underlines the impact of endorsements as mechanisms mitigating tensions that arise from diverse interactions on the platform (Wareham *et al.*, 2014). These findings thus imply the notion of increasing platform scale and scope through use of endorsement that signal how and where interaction with the platform is preferred. On a general level, this line of reasoning is consistent with modularity and platforms literature in which platform interaction as a function of standardisation is highlighted as a core tenet (Baldwin and Clark, 1997; Gawer and Cusumano, 2014; Tiwana *et al.*, 2010). Accommodating third party interests also speaks to the stimulation of platform generativity through the establishment of arm’s-length relationships with external actors (e.g., Eaton *et al.*, 2015; Karhu *et al.*, 2018).

4.7.2. Unsuccessful signals of endorsement

Our second set of findings relates to the two unsuccessful endorsement signals. Again, drawing from the language of signalling theory, failure to identify the right quality or communicate a signal helps in explaining the results in our analysis.

In the case of the unsuccessful endorsements to ‘Balance Market Demand’, the signal sought to explicate that the platform desires to increase demand in a way that is balanced across participant audiences. The endorsements did so by delineating alternatives in the usage of tags in a way that maximises the potential audience of platform participants. From our content analysis, we know that most of these endorsements had to prioritise a tag alternative over one or several others. A possible interpretation of our analysis results is offered by the concept of signal strength. While signalling areas where interaction with the platform is preferred is the foundation upon which further interaction can occur, signals such as endorsements need to provide sufficient strength to distinguish the desired quality that is intended to be signalled (Connelly *et al.*, 2010). Failure to do so can result in signals getting lost in noise and degenerate completely (Boudreau and Jeppesen, 2015).

It is plausible that an endorsement’s signalling function is contingent on a critical upper bound of possible alternative messages beyond which the signal seizes its effect. The signal is hence lost in the noise of abundant existing alternatives. Indeed, scholars drawing from signalling in their theorising of platform marketplace dynamics have noted how overcrowding on platforms can lead to “confusion, dissonance, or uncertainty” if signals cannot be decoded amidst noise (Boudreau and Jeppesen, 2015, p. 1765). In the case of OpenStreetMap, the usage of tags with this endorsement grew as the platform operator signals where, given several options, the installed base is preferred to focus attention. As the endorsement exercises authority, platform participants are forced to adapt, and in the long run, endorsements with this motive still provide sufficient certainty for the platform to grow in scale (Chintakanada and McIntyre, 2014). However, the signal does not possess necessary strength to stimulate increases in platform scope. As others have argued, while the presence of a growing installed base is a positive stimulus, efforts required for a deep and sustained adoption

of novelty might be hampered if signals are uncertain (see Chintakanada and McIntyre, 2014).

Similarly, ‘signal consistency’ (Connelly *et al.*, 2010) offers a possible explanation for the ineffectiveness of endorsements that ‘Ratify Emergent Use’. The degenerating effect discussed in relation to signal strength above is even exacerbated when signals are inconsistent. Defined as “the agreement between multiple signals from one source” (Connelly *et al.*, 2011), signal consistency is crucial for endorsements by the platform operator. In the framework of signalling, information exchanges are contingent on the consistency or “the extent to which the signal corresponds with the sought-after quality of the signaller”. This alludes to a signal’s purpose of mobilising receivers’ behaviour in a desired way. But the signalling process is ineffective “if the receiver is not looking for the signal or does not know what to look for” (Connelly *et al.*, 2010). In the case of endorsements that signal the ratification of an emergent use case, the desired effect has already happened. This presents a problem for platform operators. Signals risk appearing inconsistent if the desired quality can exist both in the future and the past. In cases where the desired effect has already happened, a signal, no matter how well intended, is pointless. The effects of tag endorsements that we surmise here have already occurred. Signalling areas of further growth the same way as with the other motives for endorsements does simply not work. This is not to say that the tags are unsuccessful. This is to say that the signal is useless as the desired growth in scale and scope has already happened, yet it did so in the absence of guidance by the platform operator. In the case of geo-data this plausibly explains the subpar spread of such tags in scale as well as the mediocre scope as reflected in the level of detail of such tags. Over time, endorsements with that motive become increasingly useless as misleading signals are learnt to be ignored by the receiving platform participants (Connelly *et al.*, 2011).

In summary, endorsements generally lower the effort thresholds needed for platform participants to engage with the platform (Baldwin and Clark, 1997; Tiwana *et al.*, 2010). Endorsement signal explicit design choices motivated by an interest in guiding desirable variation on the platform (Wareham *et al.* 2014). In case an endorsement prioritises one kind of interacting with the platform over another, external actors are forced to adapt. In the long run, endorsements increase reliability and decrease

uncertainty, leading to favourable conditions for a platform to grow (cf. Tiwana, 2014).

However, reinforcing both scale and scope of the platform through endorsements has its limitations. Using the language of signalling, we surmise that the signalled quality needs to be correctly identified and communicated contingent on signal strength and consistency. Given the correct identification of desirable interaction (the signalled quality) and signalling accordingly, endorsements are crucial for generativity. Explicating what interactions are favoured on the platform increases both scale and scope of interactions on the platform.

4.8. IMPLICATIONS

Our study comes with two major implications for research on generativity and digital platforms.

First, previous literature has demonstrated that a tension between openness and control complicates the stimulation of generative change on digital platforms (Tilson *et al.*, 2010). Simply put, if platform operators apply an ‘anything goes’ approach, contributions by platform participants might well bring about generative developments. Yet, these might not always align with the interests of the platform operator. This has paved the way for substantial contributions (e.g., Eaton *et al.*, 2015), yet crucial questions about platform generativity remain hitherto unanswered. Here, our study contributes a highly detailed and theoretically grounded account of generative change on digital platforms. Inspired by the tensions between control and openness in the literature (Hanseth and Lyytinen, 2010; Tilson *et al.*, 2010), this study is the first one to acknowledge that not all generativity is unequivocally positive. With the examination of endorsements as deliberate actions by the platform operator to stimulate generativity, we highlight how such changes can be directed to areas where it is desired.

Informed by signalling theory, we highlight how endorsement are crucial actions available for the platform operator. While this does not equate control, it makes generativity manageable to a degree. This presents a potentially powerful lever for platform operators to steer interaction across the platform. As such, our findings

underline the impact of purposeful actions steering the tensions that arise from diverse interactions on the platform and provide evidence of the ability to increase scale and scope of a platform through the use of endorsements. Here, our study speaks to how platform participant behaviour can be steered towards desirable variation on digital platforms (cf. Wareham *et al.*, 2014). It also indicates how stability induced through endorsements enable variation somewhere else on the platform, implying that areas without endorsements reduce "undesirable variance" described by others (Wareham *et al.*, 2014). Exploring these dynamics further strikes us as a fruitful line of inquiry for future platform research. For instance, we were able to study but a single kind of endorsement. The acceptance of a metadata tag is akin to the standardisation of a resource on the platform boundary (e.g., Eaton *et al.*, 2015; Karhu *et al.*, 2018). Yet, as we have pointed out, endorsement can come in a variety of actions by the platform operator. They can be explicit (Parker *et al.*, 2017) or implicit (Förderer *et al.*, 2018), perhaps even unintended. In striving to understand how generative growth and innovation in platforms can be cultivated through such actions, a range of endorsements are likely relevant and their investigation motivates future work.

Second, our study is the first systematic study and categorisation of strategic motives of endorsements. Each motive aspires to stimulate certain types of behaviour on the platform that the platform operator wishes to see implemented. As signals of otherwise unobservable qualities, platform operators stand to gain from making their design choices, strategic decisions, and future aspirations explicit and known to platform participants by using signals. Our hypothesis development approach draws from qualitative evidence and prior studies in the platform literature. As such, we are confident about the validity of the categories we identified. We are certain that the motives for endorsements are relevant for the platform literature and with a thorough reflection in the literature have the potential to generalise well across contexts. The theoretical grounding of the endorsement motives and their effects presents a meaningful implication for platform scholars. However, it is perfectly plausible that other motives exist that we have not identified. Either we did not observe them in the available empirical material or did not have enough information to motivate the formulation of an additional category (compare Kelle, 2015). Therefore, careful reflection and potential elaboration is needed in further studies on these, and other, strategic motives of platform operators to endorse (or abstain from doing so). For instance, despite obvious advantages in terms of access and data collection, our choice

of empirical setting opens a pathway to study strategic motives on platforms with different forms of ownership, governance, and design in order to understand what makes endorsements motives effective across different kinds of platforms.

4.9. CONCLUSION

In this paper we extend the existing literature with a view on how generativity on digital platforms can be stimulated by the platform operator using endorsements. Endorsements follow strategic motives and signal where participant contributions are desired on the platform. We argue that given the correct identification of desirable qualities, endorsements are crucial for generativity as they can increase the output of a platform in scale and scope. In order to increase the efficacy of such actions, platform operators should pay attention to the consistency of the sought after quality and the communication of the signal that ought to bring the desired effect about. If done right, endorsement signals provide a powerful lever for platform operators to steer generative change on digital platforms.

APPENDIX TO CHAPTER 4

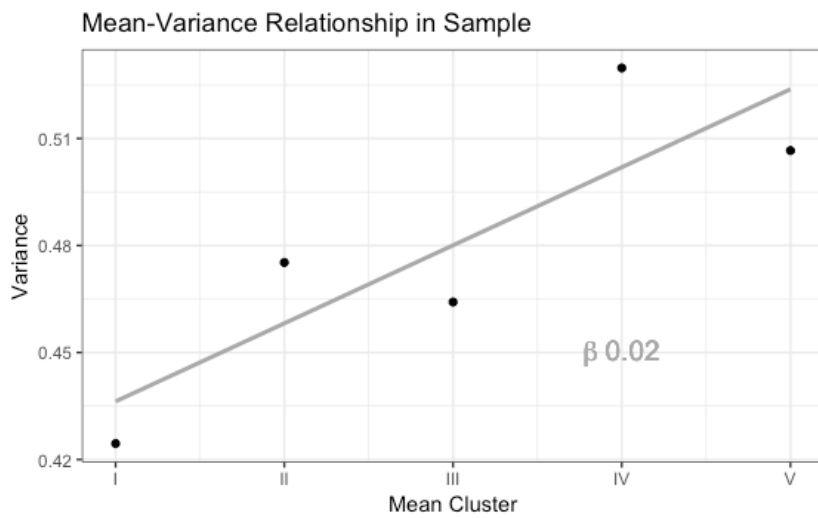
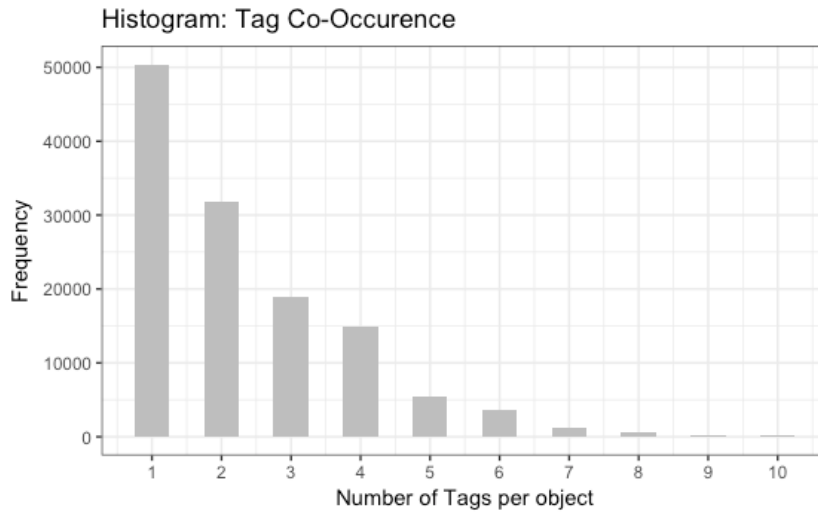
1) Full Results Chi-Square Test of Independence

Expected values in brackets - these are the values that would be expected if all endorsed tags would have the same chance of outgrowing the platform average as unendorsed tags do (18%).

Endorsement Motive	Timing	#Above Average Growth	#Below Average Growth	Chi-Square (p-value)	Sum
Commit to New Market	Before	162 (42.3)	73 (192.7)	413.08 (p <0.001)	235
Commit to New Market	After	65 (22.3)	59 (101.7)	99.52 (p <0.001)	124
Accommodate 3rd Party	Before	22 (7.38)	19 (33.6)	35.32 (p <0.001)	41
Accommodate 3rd Party	After	10 (4.1)	13 (18.9)	10.11 (p <0.001)	23
Balance Market Demand	Before	20 (8.82)	29 (40.2)	17.28 (p <0.001)	49
Balance Market Demand	After	8 (4.9)	19 (22.1)	2.47 (p = 0.057)	27
Ratify Emergent Use	Before	12 (7.92)	32 (36.1)	2.56 (p = 0.054)	44
Ratify Emergent Use	After	2 (4.1)	21 (18.9)	1.34 (p = 0.877)	23

2) Tag Counts: Histogram and Mean-Variance relationship

A linear mean-variance relationship is a key indicator of the applicability of Quasi-Poisson models for count model analysis. Following Ver Hoef and Boveng (2007) we used mean squared errors for this diagnostic step.



3) Post-Hoc Robustness Checks for Endogeneity

We used both, control function (CF) and instrumental variable (IV) approaches to address concerns of endogeneity in the reported models.

- a) For the logistic regression we used the fact that the linear probability model yielded consistent results to address the endogeneity of the of the binary treatment (i.e. tag endorsement).
 - i) We first followed a two-stage control function approach formulated by Rivers and Vuong (1988) and well documented by Train (2009, chapter 13). We modelled the residuals of the potentially endogenous treatment ('endorsement') as an additional predictor for the logistic regression and achieved consistent results and an insignificant effect for the endogenous portion as a predictor ($p = 0.551542$)
 - ii) The logistic model results are robust when modelled as a linear probability model (using an OLS estimator). Using this specification, we modelled the control variable 'Tags Documented' as an instrument in a 2 stage least squares approach (compare Wooldridge 2002, p. 85), since we assume that it does not directly influence the outcome (the coefficient of the control is not significant) but we can reasonably expect it affects the likelihood of the event 'tag endorsement' to occur. This is due to a number of endorsements occurring in bulk as well as in succession within one tag group ($key=$). We used the Wu-Hausman test for this step and can reject the alternative hypothesis that the IV model is inconsistent with the linear probability model ($H_0: IV = LPM$, $df = 359$, $p = 0.6267$). The binary treatment is hence not significantly endogenous.
- b) For the count data regression, we used an approach similar to 1a) and confirmed that the endogenous portion of the treatment (i.e. 'tag endorsement') is an insignificant predictor in the full count model ($p = 0.46885$) (see Wooldridge 2002, p. 663)

References

- Train, K. (2009), Chapter 13 in *Discrete Choice Methods with Simulation*, Cambridge University Press, Second edition
- Rivers, D., Vuong, Q. (1988), Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models, *Journal of Econometrics*, 39, pp. 347 - 366
- Wooldridge, J., (2002), *Introduction to Econometrics - A Modern Approach*, Cengage Learning, 6th Edition

4) Descriptive Statistics

Descriptive Statistics (Dataset 1)								
Variable	Min	Max	Mean	SD	1	2	3	4
1 Tag Usage Growth Rate	-0.98	28.90	0.18	0.38	1.00			
2 Objects in Database	9.38e8	4.41e9	2.17e9	1.09e9	0.06	1.00		
3 Tag Tenure	0.00	68.00	0.12	5.60	0.04	0.15	1.00	
4 Tags Documented	0.00	0.13	0.03	0.02	0.07	0.20	0.07	1.00

Descriptive Statistics (Dataset 2)								
Variable	Min	Max	Mean	SD	1	2	3	4
1 Tag Occurrence	1.000	24.000	2.400	4.572	1.00			
2 Objects in Database	1.05e7	4.41e9	2.76e9	1.17e9	0.10	1.00		
3 Tag Tenure	0.000	68.000	0.120	2.290	0.01	0.14	1.00	
4 Tags Documented ³²	0.000	0.120	0.020	0.004	0.01	-0.03	0.80	1.00

³² Variable split into quartiles to avoid multicollinearity

CHAPTER 5 – COMPLEXITY, EXTERNAL DEPENDENCE, AND THE PACE OF CHANGE IN UNBOUNDED DIGITAL PLATFORM ECOSYSTEMS

ABSTRACT

Modularity is a foundation for change in platform ecosystems. It supports interaction between loosely-coupled modules by hiding complexity within and minimizing dependence among them. However, classical modularity does not accommodate for platform ecosystems where modules are open for novel and unanticipated interactions. Common in the context of digital platforms, such “unbounded platform ecosystems” warrant attention for their specific qualities to incorporate change. Studying the open data platform, OpenStreetMap, and its ecosystem, we investigate the balance between structural complexity and external dependence. Contrary to conventional views on dependencies and change in modularity, we find that dependencies across the ecosystem increase the pace of change in core modules despite high structural complexity. Platform operators should thus manage unboundedness to attain the ability to rapidly change the platform.

A version of this chapter has been submitted for publication and is currently under review as:

Hukal, P., Henfridsson, O., Yoo, Y. (2018), COMPLEXITY, EXTERNAL DEPENDENCE, AND THE PACE OF CHANGE IN UNBOUNDED PLATFORM ECOSYSTEMS

COMPLEXITY, EXTERNAL DEPENDENCE, AND THE PACE OF CHANGE IN UNBOUNDED PLATFORM ECOSYSTEMS

5.1. INTRODUCTION

Modularity is at the core of technology ecosystems that facilitate dynamic interaction among otherwise autonomous organisations (Adner, 2017; Jacobides *et al.*, 2018; Kapoor and Agarwal, 2017). Unsurprisingly, modularity has also substantially informed prior research on digital platform ecosystems where it serves as a powerful explanation for change by reducing coordination costs (Adner, 2017; Baldwin and Clark, 2000; Schilling, 2000). However, in digital platform ecosystems, modularity alone does not exhaustively explain why dependencies across technical and organisational boundaries do not impede a platform's ability to change. In order to foster interaction in technology ecosystems, modularity would dictate to minimize interdependencies among modules. Yet, in digital platform ecosystems modules regularly draw in functionality from elsewhere without slowing down the pace of change.

In such digital platform ecosystems, a common strategy is to extend a platform's stable codebase by interacting with external add-on modules, typically developed by third-party developers (de Reuver *et al.*, 2018; Tiwana *et al.*, 2010; Yoo *et al.*, 2010). The use of shared artifacts (e.g., Software Development Kits), interfaces (e.g., Application Programming Interfaces), or standards and rules allows for novel and often unanticipated interactions across architectural boundaries (Eaton *et al.*, 2015; Parker *et al.*, 2017).

Indeed, these interactions across boundaries are an important driver of digital platforms' ability to utilize change and innovation (Henfridsson *et al.*, 2018; Yoo *et al.*, 2010). In this paper, we argue that such consequences of unboundedness are unaccounted for in the extant platform ecosystem literature, and that an increased understanding of its implications is essential for making well-informed strategic decisions in platform ecosystems.

In particular, we argue that digital platform ecosystems differ from traditional modular systems in crucial respects: digital platform ecosystems are unbounded systems. While

digital platform modules and the code they are comprised of are doubtlessly modular (Baldwin and Clark 2006; de Reuver *et al.* 2016), the prevalent view on modularity in studies on platform ecosystems risks misrepresenting important qualities of digital technology: Modules on digital platforms are *system-agnostic*, *reprogrammable*, and *interactively designed*.

First, digital platform modules relax the ideal of a typical one-to-one relationship (Ulrich, 1995) between the function and the structure of a module. Instead, platform modules are often system-agnostic (Yoo *et al.*, 2010) in that they can be integrated across design hierarchies (Clark, 1985). Departing from using the hierarchy (one-to-many) as a mechanism for integrating modules, technical interoperability allows many-to-many relationships among modules.

Second, digital technology stimulates innovation in absence of central design and task structures (Baldwin and Clark, 2000; Zittrain, 2006). Modules in digital platform ecosystems consist of software that is reprogrammable and editable (Kallinikos *et al.*, 2013b; Yoo *et al.*, 2010). This invites platform operators to appropriate modules shaped by unanticipated developments rather than by central design implementation (Boland *et al.*, 2007).

Third, digital platforms modules are often developed through interaction with heterogeneous and distributed actors. As a result, the locus of innovation transcends the organisational boundaries of the platform operator (Parker *et al.*, 2017).

However, despite the emergence of a body of literature that traces unique aspects of digital technology (Kallinikos *et al.*, 2013b; Yoo *et al.*, 2010), we know little about the effect of unboundedness on a modular system's ability to change. Yet, winner-take-all dynamics in platform competition (Parker *et al.*, 2017), require platform operators to mobilize the appropriate organisational responses when change is urgent.

In this paper, we report an empirical study of the geo-data platform OpenStreetMap designed to test how unboundedness affects the pace of change of core platform modules. In response to dynamics in its ecosystem, OpenStreetMap often needs to change rapidly. For instance, during one episode, the platform could no longer draw

from an external technology ‘TileMill’³³ – an open source framework to visualize and interact with mapping data. In 2015, TileMill’s parent organisation, Mapbox, discontinued active development on the project and instead marketed a proprietary solution with different technical specifications. The OpenStreetMap platform hence had to react quickly and needed to adjust the architecture of its core modules in order to provide the functionality expected by complementors so as to not risk them migrating to other platforms.

Through our analysis, we will argue that organizing in digital platform ecosystems differs from traditional modular systems in at least one fundamental aspect. Traditional modular systems seek to contain complexity through information hiding, decoupling, and encapsulation (Langlois, 2002) to decrease coordination costs (Tiwana, 2008). In contrast, digital ecosystems – platforms, the modules that extend them, and the actors using them – promote complex and open-ended interactions among diverse technologies developed by autonomous organisations to facilitate change.

It is important for organisation scholars to understand the dynamics of interactions in digital platform ecosystems. We investigate these dynamics in the context of OpenStreetMap and explore the role of dependencies across the ecosystem by analysing the changes to platform module source code. Our results indicate a delicate balance in the design of platform core modules. Particularly, we find that platform modules whose structural complexity is to a large extent determined by external dependencies incorporate change at a higher pace than other designs. We make important contributions to the extant literature on platform ecosystems by (1) conceptualizing unboundedness of digital platforms for organisation research and (2) formulating a strategic option for platform operators that embraces unboundedness as an organizing principle in digital platform ecosystems in order to foster change.

5.2. CONCEPTUALISING UNBOUNDED PLATFORM ECOSYSTEMS

5.2.1. Unboundedness of Digital Platform Ecosystems

³³ By now the software has been launched again as an open source project; <https://tilemill-project.github.io/tilemill/>

Most literature in strategy and technology innovation management deals with modular product systems that are - at least in parts - based on physical components (Adner and Kapoor, 2010; Ethiraj, 2007; Hannah and Eisenhardt, 2018). Accordingly, modules tend to be seen as components with fixed boundaries within which resources are bundled to fulfil predefined tasks (Simon, 1962, 2000). The final product system is made up by the entirety of modules required to achieve a set of desired functionalities (see e.g., Ulrich, 1995).

Technology ecosystems are regarded as aggregations of modular product systems and just as any other modular system, are subject to the same risk of jeopardizing performance through complexity. In this vein, modularity serves as the means to contain complexity that arises from mapping nested components to functions within self-contained systems (Baldwin *et al.*, 2014). As managerial effort is required to resolve such interdependencies, minimizing complexity in modular designs is vital to avoid slowdowns in maintenance and innovation. To ensure the ability to change within ecosystems, it is hence important to avoid complex and unforeseeable module interactions.

As a result, resolving module interdependencies is paramount in bounded modular systems in order to reduce complexity. This usually involves substantial effort on the part of the platform operator as every module within a product system has to be integrated into the system's design hierarchy. Brusoni and Prencipe (2013), for instance, demonstrate how continuous managerial control of interfaces is required to retain the balance between 'responsive and distinct' modular designs within a system. Similarly, Ethiraj (2007) found increased efforts goes into modules that present critical interdependencies across organisations in the PC industry ecosystem.

The economics of ecosystems thereby constrain strategic options for firms. This is especially apparent when bottlenecks arise in product systems. Hannah and Eisenhardt (2018), study competition and cooperation around components that are loci for interdependencies in an ecosystem. They observe that strategic decisions, such as which product-market segment to address, are determined by key players occupying points of crucial intersections in the ecosystem (Hannah and Eisenhardt, 2018). In some cases, this forces organisations to extreme measures. For instance, while seemingly counterintuitive to the dogma of decoupling – even vertical integration can

remain as the only adequate response for firms if it serves the goal of resolving critical interdependencies between firms (see Adner and Kapoor, 2010).

In contrast to conventional views of modularity, digital platform ecosystems embrace complexity as a driver of innovation. Therefore, the goal of digital platform ecosystems is to promote complex and dynamic interactions between affiliated organisations in a way that gives rise to serendipitous interactions (Yoo *et al.*, 2010; Yoo *et al.*, 2012). Although digital platform ecosystems are built on the principle of modularity, modularity in digital platform ecosystems does not reliably reduce the complexity among modules (Kapoor and Agarwal, 2017). As such, understanding digital platform ecosystems requires a shift in thinking about the role of complexity in modular architectures of digital technologies.

Digital technology marks a departure from the nested and fixed product architectures prevalent in most modular product systems. Digital technology is characterized by layered-modular architectures in which components comprise software (Yoo *et al.*, 2010). Rather than designing, providing, and maintaining all functionality within the boundary of a module, digital technology allows the orchestration of functionalities by managing openness and interoperability across a platform's ecosystem (Boudreau, 2010; Kapoor and Agarwal, 2017; Parker *et al.*, 2017). We refer to this as unboundedness of digital platform ecosystems.

5.2.2. Modules in Unboundedness Ecosystems

Unboundedness is essentially a strategic decision by the platform operator to promote architectural designs that draw on unique attributes of digital technology to include modules that are *system-agnostic*, *reprogrammable*, and *interactively designed*.

First, platform modules are system-agnostic (Yoo *et al.*, 2010). Creating digital platform ecosystems does not follow a central plan that implements a preconceived design and task structure (Baldwin and Clark, 2000). Of course, a platform is often centrally designed by the platform operator (cf. Wareham *et al.*, 2014). However, an ecosystem surrounding the platform typically grows organically. Indeed, the virtue of digital platforms is that their architecture allows the introduction of novelty through unprompted changes (de Reuver *et al.*, 2018, Wareham *et al.*, 2014). As these changes

are often initiated by third-party developers, novel developments transgress the boundary of the platform (de Reuver *et al.*, 2018; Yoo *et al.*, 2010; Zittrain, 2006). Rather than focusing on pre-existing design rules (Baldwin and Clark, 2000), the premier condition for integration is technical interoperability among modules provided via an open interface. Since this interoperability substitutes managerial coordination to resolve module interdependencies, relaxes the extent to which a modular system is constrained by interdependencies among modules. This makes modules in digital platform ecosystems agnostic to the overall system (Yoo *et al.*, 2010).

Second, the form and function of digital platform modules can be swiftly unbundled. Should a module have to be changed, its program code can be easily adjusted to integrate with the system to which it contributes (Yoo *et al.*, 2010). As a result, modules in digital platform ecosystems are malleable to changes after the fact by means of reprogramming (cf. Kallinikos *et al.*, 2013b). Past work has highlighted how software modules enable changes to the designs of modular systems due to their reprogrammability even after their implementation (Lee and Berente, 2012).

Third, modules in digital platform ecosystems are interactively designed. The participative architecture of source code renders modules editable through cumulative interaction (Baldwin and Clark, 2006). Modules are often developed through shared efforts by heterogeneous and distributed contributors. Digital platform ecosystems thus shift the locus of interaction and innovation beyond the organisations that created the focal platform in the first place (Parker *et al.*, 2017). Boudreau (2010), for instance, evidences the relationship between innovative activity in platform ecosystems and the degree to which platforms open up modular components for reuse (Boudreau, 2010). This also affects the long-term evolution of digital platform ecosystems. In their study of the Apple iOS ecosystem, Eaton *et al.* (2015), demonstrate how the platform evolves beyond the original design as modules at the platform boundary are shaped by heterogeneous and distributed actors - generating results largely outside of the control of Apple, the platform operator.

In summary, design, integration, and recombination of modules in digital platform ecosystems are largely unconstrained since software can easily be altered. The attributes of software modules allow modules to be integrated irrespective of their

position in a design hierarchy. In addition, modules are interactively designed in a bottom-up fashion, sourcing software code from various origins in an ecosystem of digital technologies and igniting unanticipated developments (Yoo *et al.*, 2010). As all these dynamics regularly unfold across organisational and architectural boundaries, digital platform ecosystems become unbounded systems.

5.2.3. Organising Unbounded Platform Ecosystems

The way we conceive the unboundedness of platform ecosystems poses a challenge for platform operators. Indeed, central to the discussion on platform ecosystems has been the question of how a platform operator can create a vibrant ecosystem of heterogeneous complementors while remaining in control (e.g., Boudreau, 2010; Wareham *et al.*, 2014). So far, scholars have mostly focused on structural arrangements in order to manage tensions between the platform operator who wants to retain control and heterogeneous complementors who want to pursue their own ideas. For instance, attention was given to the design of components on the platform boundary (often referred to as “boundary resources”) that render interaction with the platform (Eaton *et al.*, 2015; Ghazawneh and Henfridsson, 2013), the search for regulatory regimes with mutual benefits (Parker and Van Alstyne, 2017), or the extension of base functionality of the platform (Förderer *et al.*, 2018).

In this paper, we advocate an additional mechanism. That is, a platform operator who wants to remain responsive to changes in the platform’s ecosystem, can embrace the unboundedness of digital technology and design platform modules accordingly. The reduced need for *a priori* top-down coordination for the integration of modules holds a promise for platform operators. As a result of digital technology module attributes (agnostic, reprogrammable, and interactive), digital platform ecosystems are less constrained by interdependencies among technologies when compared to traditional modular systems. To the contrary, complexity in the form of couplings between modules enables interaction between otherwise unrelated organisations governed almost exclusively by technical interoperability (Jacobides *et al.*, 2018; McIntyre and Srinivasan, 2017). Since interoperability substitutes managerial control, the extent to which unbounded systems risk jeopardizing performance through resolving module interdependencies is limited. Growingly complex interdependencies among modules enable a platform’s participation in a shared architecture. This in turn can be a driver

of unprompted change coming from outside the platform itself (Parker *et al.*, 2017; Parker and Van Alstyne, 2017). The platform operators' control over the platform therefore involves the cultivation of the platform's ability to change and accepting a lack of direct influence over the entire unbounded architecture.

As such, embracing the growing complexity of an unbounded digital platform ecosystem, a platform operator can attain malleability for the platform core i.e. the ability to incorporate change rapidly (Agarwal and Tiwana, 2015; Tiwana *et al.*, 2010). The ability to change parts of the platform at a high pace is highly desirable for platform operators. Malleability enables the platform operator to rapidly respond to unanticipated changes by incorporating developments from its ecosystem. A high pace of change translates into resilience to react to external shocks (Olleros, 2008), as well as adaptability to multiple contexts (Spagnoletti, Resca, and Lee, 2015). Ultimately, a platform's capacity to absorb change is related to its ability to respond to unanticipated changes brought about by heterogeneous third-party developers (Eaton *et al.*, 2015). While the core promise of modularity is to increase the ability of a technical system to change, the received view of modularity suggests that platform operators need to adhere to modular design principles, most importantly: minimizing the dependencies between modules (Tiwana, 2015). However, as platform ecosystems become unbounded, platforms embrace complex interactions as the source for serendipitous innovation. This implies that the pace of change of platform's core modules can be influenced by the extent to which the platform operator manages unbounded interactions between platform modules and the wider platform ecosystem.

5.3. THEORY DEVELOPMENT AND HYPOTHESES

As a design choice, unboundedness is a function of the architecture of the underlying technology modules of a platform. For the purpose of this research, we distinguish module architecture along two dimensions: *structural complexity* and *external dependence*. By structural complexity, we refer to the overall internal structure of a module. This can be understood by the number of interconnections within a module (Baldwin and Clark, 2000; Tiwana, 2014, p.78). By external dependence, we refer to the extent to which the boundary of a module is permeated by interdependencies with modules from the platform's ecosystem.

Conceptually separating the two dimensions along attributes of low or high structural complexity and low or high external dependence yields four configurations of modules in platform ecosystems. Each of these four configurations represents a design pattern that is available to platform operators. In line with our theoretical development we argue that these design patterns are associated with different degree of unboundedness, and thus are likely to differ in their pace of change (compare figure 1 below).

Modules with low structural complexity and low external dependence represent the classic design pattern of modularity with loose coupling (Baldwin and Clark, 2000, p. 65). Loose coupling arguably increases a system's overall ability to change (Simon, 1962). The advantages of modules with low structural complexity and low external dependence as a strategic option for technology design and management are undisputed (Schilling, 2000). Such modular designs free organisational and technical change processes of inertia thus increasing the ability for the encompassing system to change (Sanchez and Mahoney, 1996). Indeed, in the domain of digital technology, recombinant innovation is posited as a direct result of the modularisation of software code (Baldwin and Clark, 2006).

Consistent with this argument from traditional modularity, we hence expect a high pace of change for modules with low structural complexity and low external dependence. We refer to such modules as "classical modules" and suspect they are likely associated with higher pace of change than modules with low structural complexity but with high external dependence. Extensive couplings across modules violate modular design principles and are likely to slow down change due to high maintenance (Simon, 1962). Thus, we hypothesize:

H1: Compared to classical modules, modules with high structural complexity and low external dependence are associated with a low pace of change.

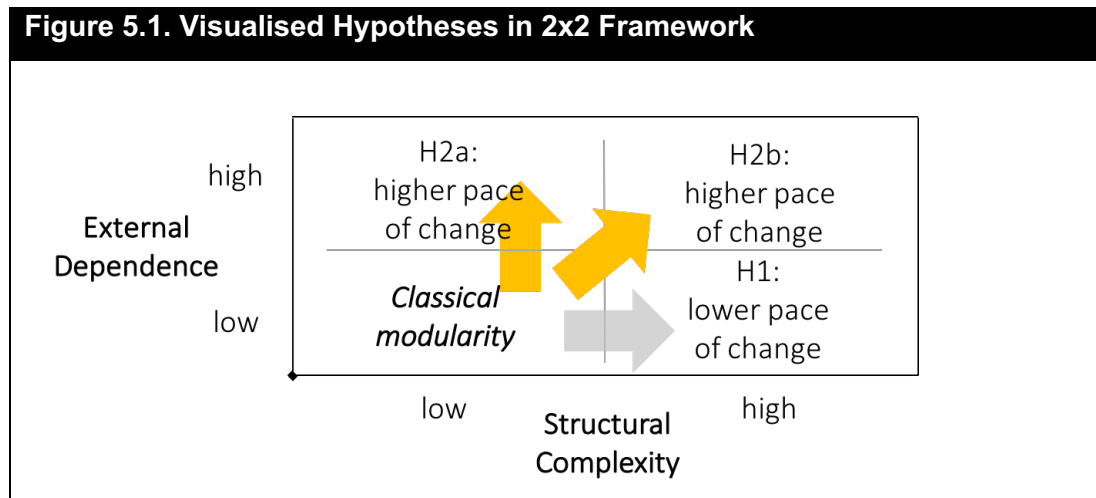
As we argue above, digital platform operators have the choice to design modules in a way that differs from classical modularity. While clearly not all modules on a digital platform abandon established modular design principles, digital platform operators are likely to leverage the attributes of digital technology to “invert” the firm’s innovation through third-party complementors. (Parker *et al.*, 2017). This diverges from the loose coupling expected in classical modularity (cf. Tiwana, forthcoming). Indeed, by what has been called ‘mirror breaking’ (Colfer and Baldwin, 2016) describe that architectural designs in technology ecosystems deviate from development trajectories expected by modularity (see also Constantinides *et al.*, 2018). Complex interactions are embraced on digital platforms as functionality can be derived from interdependencies across the ecosystem even if these couplings induce structural complexity. Therefore, we expect that modules with high numbers of external dependencies across a digital platform ecosystem are likely to have a higher pace of change than classic modules with low structural complexity and few external dependencies. Thus, we hypothesize:

H2a: Compared to classical modules, modules with low structural complexity and high external dependence are associated with a high pace of change.

H2b: Compared to classical modules, modules with high structural complexity and high external dependence are associated with a high pace of change.

In summary, we argue that unlike traditional complex modular systems consisting of physical components, classic module designs in a digital platform ecosystems may not always be the most beneficial for a platform’s ability to change. Specifically, with the platform operators’ ability to invert the innovation with third-party developers (Parker *et al.*, 2017), leveraging unique properties of digital technology (agnostic, reprogrammable, and interactive) (Yoo *et al.*, 2010), we expect that modules deviating from classic modules are likely to show higher pace of change. At the same time, even modules in a platform ecosystem will pay a penalty if they deviate from the classic

modular design and if these modules do not take advantage of the inversion of the innovation. Figure 5.1 visualises our hypotheses.



We next introduce the empirical context and derive appropriate measures to test our hypotheses.

5.4. EMPIRICAL ANALYSIS

5.4.1. Context

OpenStreetMap is an open source platform that provides geo-spatial data and related capabilities to external actors (Ramm *et al.*, 2010). OpenStreetMap competes with popular proprietary web services such as GoogleMaps, TomTom, or HERE. Competition of geo-spatial data services centers around offering up-to-date location data, at the highest accuracy, across a wide range of devices and operating systems (Parsons, 2013). In this respect, OpenStreetMap often outperforms proprietary alternatives (Arsanjani *et al.*, 2015; Cipeluch *et al.*, 2010). This success is attributed to its open source character allowing for rapid incorporation of changes to data and technology by a heterogeneous and distributed group of developers (Vandecasteele and Devillers, 2015) as well as the ease of building derivative products (Amirian *et al.*, 2015).

We conceptualize the geo-spatial database as the technology foundation of the platform; a set of relatively stable functionalities with low variety that acts as the base of the digital platform (Yoo *et al.*, 2010). The platform consists of a simple set of functionalities aimed at providing vast amount of geo-spatial data in the form of

locations and meta-data descriptions (Ramm, 2015). This provides the basis for component reuse through interaction with platform modules that add functionality (Baldwin and Woodard, 2009). The platform modules are add-on software projects that extend the database functionality of the core. In total, eighteen core modules are available to complementors, such as external application developers, wanting to use geo-spatial data handling capabilities in their own offerings.

5.4.2. Data

OpenStreetMap coordinates development work on GitHub which makes the source code of all core software modules openly accessible. GitHub is a public version control and management system popular among open source developers (Dabbish *et al.*, 2012). Using the GitHub web API, we downloaded all source code changes made to OpenStreetMap platform core modules between 2007 and 2017. GitHub's version control features include metadata such as timestamps, affected source code file, and changed lines of code with every change log. This allows for a detailed reconstruction of the changes made to the platform modules over time.

Some 59,000 source code changes have been committed to the 18 core modules on the platform. We saved every change in raw text format in a separate file to create distinct code versions for each respective module. Using a simple text analysis approach based on regular expressions, we extracted executable commands in the source code files to infer properties of each module at any given point in time.

5.4.3. Analytical Approach

Our unit of analysis is a single software module. As such, we focus on pieces of software containing add-on functionality and the ability to incorporate change. Specifically, we explore the effect of structural complexity and external dependence on the pace of change of a module. Choosing the analysis on the level of the individual module services two purposes. First, modules represent parts of the platform that bundle functionality and can be used to create derivative products based on provisions in the platform core (Baldwin and Woodard, 2009). As such, modules are system components and include all tasks and decisions required in designing, operating, and maintaining the platform (Colfer and Baldwin, 2016). Second, the focus on modules allows to make claims about the evolution of the platform as a whole in so far as the

ability to change a module reflects the ability to adapt core processes of the platform (cf. Baldwin and Woodard, 2009).

5.4.4. Measures

All main measures used in the analysis are based on two pieces of information derived from the source code analysis. First, the number of dependencies in a source code file. Second, the number of functions provided in the code base of a module. Based on the number of dependencies and the number of functions we compute measures of structural complexity and external dependence of the platform modules. As all changes are time stamped we are able to create a data set of changes made to module architecture over time. See the technical note in the appendix for a detailed description.

Dependent Variable: Pace of Change

Our dependent variable, *ChangeTime*, is the time interval between two changes made to module functionality. That is, we capture the time (in minutes) that passed between two changes that either add or eliminate functions in module source code. The rationale for including the time interval between functionality changes derives from modular systems thinking. Modularity dictates to keep the number of interdependencies between constituent system parts to a minimum. Failure to do so results in high maintenance effort and a limited ability to incorporate changes to the functionality of a module as interdependencies first need to be resolved (Fixson and Park, 2008; Lindberg *et al.*, 2016). The measure thus aligns with metrics of platform evolution as it presents an apt proxy for its ability to change (see Tiwana *et al.*, 2010). On one hand, changes to functionality mirror the evolving feature space of a module at any given point in time. On the other hand, changing functionality captures a meaningful development task aimed at adjusting module scope in response to external circumstance.

Structural Complexity

Structural complexity is a critical metric in investigation of a platform's ability to change (Tiwana, 2014; p.67). It captures the expected effort needed to incorporate changes in architectural designs (Daniel and Stewart, 2016; Woodard *et al.*, 2013). We derive an intuitive measure of structural complexity of platform modules from the raw text source code files. We refer to structural complexity simply as the ratio of all

dependencies to functions in the source code base of a platform module. Dependencies are, for instance, libraries used in the source code or couplings among modules. The number of dependencies over functions reflects a module's interconnections (Tiwana, 2014; p.78). In relation to functions, dependencies highlight the amount of additional resources needed to provide the intended functionality of a module. Deriving dependencies from the information flow within modules aligns with past metrics such as quoted patents (Ethiraj and Posen, 2013) or the coupling of modules (Baldwin *et al.*, 2014).

External Dependence

We include the source of dependencies in our investigation. Digital technology regularly draws in functionality from sources that transgress organisational boundaries (Lindberg *et al.*, 2016). As we strive to deepen our understanding of unbounded digital technology ecosystems it is crucial to qualify the source of technical interdependencies. In addition to the mere count of dependencies, we therefore also queried for the name of the dependencies used in module source code. This allows us to identify the architectural source of a dependency by drawing from technical documentation.

We treat dependencies as external if a deliberate decision was made to extend the intended functionality of a core module. Additional data sources informed the qualification of the source of dependencies as either internal or external to the platform. Mainly, we used two sources to get additional information on each platform module. One, we studied the documentation provided on GitHub. As platform modules are separate software projects, each module is documented and maintained in its own code repository. Module developers introduce their work and it considered good practice to disclose any additional resources necessary to run the code and operate a module. We then identified the dependencies named in the documentation and classified them accordingly as external or internal to the module.

We used this information to infer a module's interdependencies with other code bases, i.e., resources a module is reliant on from technologies in the wider ecosystem. Additionally, we used the guidelines and documentation offered by the used high-level programming languages such as Java, C, or Python. This helped us to infer what dependencies are standardized parts of the technology used to implement a module. A

reference to a standard library would thus be classified as an internal dependency since developers could draw from the resource ‘off-the-shelf’ as part of their runtime environment³⁴.

External interdependencies are crucial in the development of digital technology. In the context of OpenStreetMap, the use of external dependencies is especially salient when it comes to specialized knowledge of geographical information systems. Consider for instance the amount of effort each developer would have to undertake to compute accurate map projections of geo-locations in line with conventions such as the Mercator projection.

Providing coordinates that produce reliable and accurate representations of real world objects is a demanding task for it requires the implementation of complex mathematical formulae into software code. Faced with this task a module developer has two choices. One could either produce the functionality from scratch in a proprietary implementation. Alternatively, one could decide to include a pre-made solution in the form of a code library in the source code and rely on the work of others. The popular library “*PROJ.4*”, for example, is used by OpenStreetMap modules. It provides algorithms to compute and convert multiple mapping projections and is curated by a small team of subject matter experts³⁵. Using external libraries such as this avoids redundancy in development work and ensures consistent performance irrespective of the individual developer. While the approach affords a near effortless extension of the capabilities of a software module, it increases couplings across modules.

Control Variables

In addition, we include two sets of context relevant measures to control for alternative explanations. First, we control for technical attributes of the modules by including measures that qualify design changes of modules over time.

Technical Integration

We control for changes in module architecture by including a measure of technical integration of module design. We use the ratio of lines of code to source code files to

³⁴ We provide illustrative examples for the distinction between internal and external dependencies in the appendix

³⁵ See <http://proj4.org>

capture architectural integration of a module as a potential factor for a module's ability to incorporate change. Lines of code is an established metric in software engineering – often used as a post-production proxy for complexity as it represents the amount of instructions in a program (Morozoff, 2010). In relation to the number of module source code files, lines of code captures module design decisions such as refactoring or splitting source code files. The control thus accounts for the pace of change given the relative size of module parts (i.e. source code files).

Technical Debt

We also include a variable for technical debt by measuring the number of source code files per dependency in the code base. Technical debt reflects past design decisions in module development that might have an effect on present performance of a module (Woodard *et al.*, 2013). For instance, an increase in the ratio would imply the reduction of technical debt as the number of source code files grows while the number of total dependencies in the code remains stable.

Next, we control for dynamics in the development activity of individual modules.

Secondary Changes

We include a measure of secondary changes that captures the number of source code changes that immediately follow adjustments of functionality. Changes to software come in a variety of forms and not all source code changes bring about significant alterations to functionality (Chapin *et al.*, 2001). Including a measure for the number of changes that occur between adjustments of functionality thus accounts for development activity that does not directly affect functionality.

Change Rate

Finally, we control for the rate of all changes made to a module in a given week. The rationale of this variable is to ensure that the relative differences in the pace with which modules incorporate change is not merely driven by the overall frequency of changes to a module.

Tables 5.1 and 5.2 display summary statistics and variable definitions.

Table 5.1. Descriptive Statistics

Variable	Min	Max	Mean	SD	1	2	3	4	5	6	7	8
1 LN ChangeTime (all functions)	0.00	15.02	12.1	4.4	1.00							
2 LN ChangeTime (new functions)	0.00	15.04	12.2	4.6	0.93	1.00						
3 External Dependencies	0.00	1094.00	241.8	295.6	0.14	0.14	1.00					
4 External Dependencies (strict)	0.00	1345.00	311.7	361.8	0.14	0.14	0.98	1.00				
5 Technical Integration	0.66	183.00	19.7	14.9	0.12	0.12	0.19	0.18	1.00			
6 Technical Debt	0.00	242.00	10.0	22.0	0.04	0.06	0.23	0.24	0.14	1.00		
7 Secondary Changes	1.00	47.00	4.0	3.5	0.28	0.25	0.07	0.06	0.09	0.05	1.00	
8 Change Rate	1.00	18.00	1.5	0.8	0.15	0.14	0.17	0.18	0.17	0.05	0.02	1.00

*N = 1388 (1260 for alternative DV; Change Time (new functions))
all variables in non-standardised form
variables 3 & 4 not used in same model*

Table 5.2. Variable Definitions

Variable	Description
1 ChangeTime (all functions)	The time interval (measured in minutes) between two consecutive changes to a module's functionality (adding or removing functions in the source code)
2 ChangeTime (new functions)	The time interval (measured in minutes) between two consecutive changes to a module's functionality. Here only the addition of new functions is measured
3 External Dependencies	The number of references to external dependencies in the module source code.
4 External Dependencies (strict)	The number of references to external dependencies in the module source code. (all dependencies that are not part of the focal module are classified as external)
5 Technical Integration	The ratio of lines of code per source code file. Lines of code are measured by counting line breaks in source code files
6 Technical Debt	The number of source code file per dependency in the source code.
7 Secondary Changes	The total number of changes that do not affect functionality occurring between two changes to module functionality
8 Changerate	The number of changes made to a module per week

5.4.5. Hypothesis Tests

We test our hypotheses in two stages. First, we establish contrasts in the relative pace of changes across modules. Here, we are interested in relative differences between the average speed with which core modules of the platform incorporate change contingent on the broad architectural design patterns in our framework. This step is necessary to establish the contrast for later comparison as suggested in our hypotheses. We use the dimensions of structural complexity (i.e. the ratio of all dependencies to functions in the source code) and external dependence (i.e. the share of dependencies in the code base stemming from outside the module).

In the second step, we analyse the effects of external interdependencies in each of the design patterns to substantiate the role of external dependence in the pace of change. We use panel specification of the development data within each module with mixed fixed and random effects. Using mixed effects enables us to model a comparison of modules nested in design patterns in a flexible yet robust way³⁶ (Balazsi *et al.*, 2017). The fixed portion of the model uses a within group estimator that captures variation pertaining to each individual module while absorbing unobserved variation between modules (Wooldridge, 2016; p. 435). We include a random component to allow for a variable intercept of each module independent of its respective design pattern. Longitudinally, the panel is identified at a weekly periodicity. The independent variable is log transformed to avoid overestimation of extreme values. The model is denoted as:

$$\log(Y)ChangeTime_{i,t} \sim \mu_i + \beta 1 Design\ Pattern \times \beta 2 External\ Dependencies_{i,t} + Controls_{i,t} + (Module_i) + \epsilon_{i,t}$$

Where *ChangeTime* is the time interval between changes made to a module's functionality. The variables $\beta 1$ and $\beta 2$ denote the main fixed effects. $\beta 1 Design\ Pattern$ is a 4-level factor variable representing the design pattern of the module. The classical modularity case 'Low structural complexity - Low external dependence' serves as the reference group. $\beta 2 External\ Dependencies$ denotes the number of unique external

³⁶ We used the statistical programming language R. For a discussion of this implementation see Galecki A., Burzykowski, T. (2017), *Linear Mixed-Effects Models Using R - A Step-by-Step Approach*, New York:Springer

dependencies in module code. The distinction of external and internal dependencies follows the classification described above.

We include an interaction of β_1 and β_2 as our main theoretical interest is in the effect of external dependencies for modules in each design pattern. The random effect is denoted by $(Module_i)$ allowing for the intercept of each module to vary independently from the population average intercept of the design pattern.

We control model specifications by employing the following tests. One, results from Pesaran test for cross-sectional dependence indicate cross-sectional dependence assumptions are met ($z = -0.4707$, $p\text{-value} = 0.638$). While a Lagrange multiplier test indicates that fixed time effects are not needed ($\chi^2 = 0.36655$, $df = 1$, $p = 0.545$), we detected serial correlation (Breusch-Godfrey test: $\chi^2 = 62.165$, $df = 1$, $p < 0.001$). A Durbin-Watson test ($DW = 1.0413$, $p < 0.001$) indicates autocorrelation and we adjust the covariance structure to a first-order autoregressive process to account for the serial correlation accordingly.

5.6. RESULTS

We present our results in two steps following our analytical approach. First, we present our findings from the cross-module analysis, followed by the within-module panel analysis.

5.6.1. Cross-Module Analysis – Comparison

To conduct a cross module-analysis, we computed and compared the average time interval³⁷ between changes made to functionality of modules in each of the four design pattern. Each design pattern is distinguished by values below or above the median value of the respective conceptual dimension. Table 3 shows the results of the cross-module analysis. The comparison yields insights into the ability of a module to incorporate changes made to its functionality given the overall architectural design of the module. An overview of the mean values per design pattern are provided in table 5.3.

Table 5.3. Mean Values per Design Pattern

³⁷ The time intervals are log transformed and centred around the mean value across all modules.

Design Pattern	Structural Complexity	External Dependence	Change Time (all functions)
low-low	0.476	0.229	-0.1310
low-high	0.476	0.489	-0.0782
high-low	3.455	0.229	0.2490
high-high	3.455	0.489	-0.4114
<i>Median Structural Complexity (across patterns): 1.420</i>			
<i>Median External Dependence (across patterns): 0.380</i>			

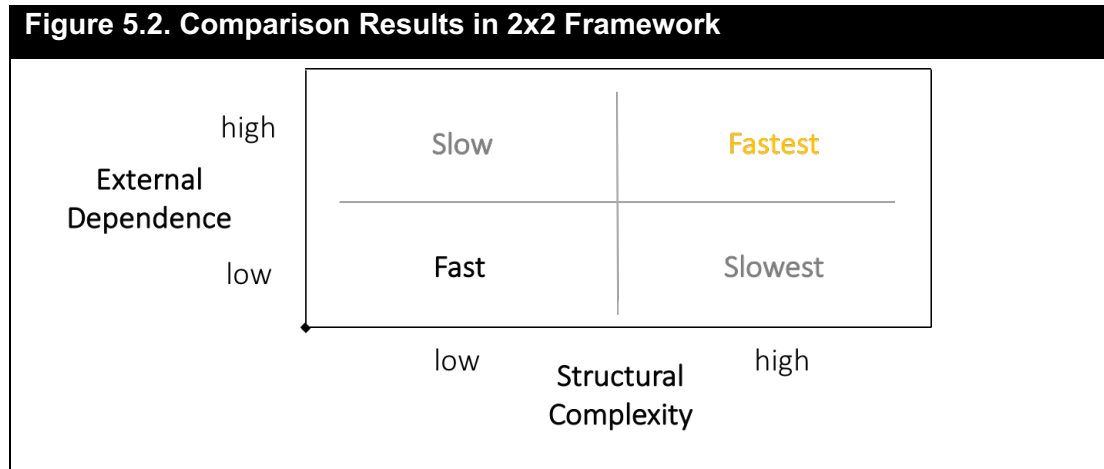
Modules with a *high structural complexity and a low share of external dependence* are – in relation to all other modules – by far the slowest when it comes to changes to their functionality (Change Time = 0.2490). This is not surprising as it speaks to the virtue of traditional modular designs. Modules with complicated designs within a fixed boundary perform relatively poorly as interdependencies need to be circumvented or resolved before functionality of a module can be adjusted. As a lot of effort goes into maintenance and coordination work, modules in this design pattern evolve slower than in other designs.

Modules with *low structural complexity and a high share of external dependencies* present a somewhat higher pace of change. Here, we observe time intervals that are close to zero (Change Time = -0.0782), indicating that the pace of change in these modules is close to the average across all modules. While this marks a slightly faster pace of change compared to the *high complexity - low dependence* design pattern, the pace of change is still the third slowest in our sample.

Next, we find that modules designed with *low structural complexity and a low share of external dependencies* develop at a relatively high pace (Change Time = -0.1310). This average time interval between changes made to their functionality makes modules in this design pattern the second fastest in the sample. This once again underlines the ideal of traditional modular design principles as the architecture of a module is kept simple with little to no dependencies across modules.

However, by far the modules that exhibit the highest pace of functionality changes are modules in the design pattern *high structural complexity and high external dependence* (Change Time = -0.4114). This indicates that these modules developed fastest in relation to all other modules despite seemingly severe deviations from

established design principles of traditional modular systems. Figure 5.2 illustrates the results of the first stage.



We used analysis of variance (ANOVA) and non-parametric two sample tests for statistical validation of the comparison across design patterns and their respective pace of change³⁸. Overall ANOVA results confirm the significance of the four design patterns on the different time intervals ($F = 10.98$, $DF = 3$, $p < 0.001$). Using the design pattern with the highest pace of change ('high complexity-high external dependence') as a reference group, the coefficients for the remaining patterns are: low-low $+0.27$, $p = 0.054$; low-high $+0.31$, $p < 0.05$; high-low $+ 0.65$, $p < 0.001$. This confirms the significance and relative position of the mean values of evolution speeds for modules in each pattern. We additionally tested the differences in time intervals using non-parametric two sample tests³⁹ to confirm the contrasts. Again using the pattern with the highest pace of change ('high complexity-high external dependence') as a reference group, the contrast with the mean of all other groups is significantly higher ($W = 81620$, $p < 0.05$). Likewise, the two fastest groups ('high complexity-high external dependence') AND ('low complexity-low external dependence') are significantly faster than the remaining groups ($W = 186660$, $p < 0.001$).

In sum, the findings from the comparison of the broad design patterns reveal differences in the performance of modules contingent on four broad architectural designs. Specifically, we find that modules designed in line with classical modularity

³⁸ We imputed values below 2.5% and above 97.5% percentiles with values rounded to the nearest digit. This does not affect the clustering results, but serves robustness for the ANOVA as it ensures residual normality.

³⁹ Wilcoxon Rank-Sum tests with alternative hypothesis true mean is smaller than contrast

change faster than modules that violate core modularity principles. However, modules whose structural complexity is to a large extent determined by external dependencies seem to outperform modules in all other design patterns in our sample.

While these findings are interesting in itself, the question remains whether the influence of external interdependencies is responsible for increases in the speed in which modules evolve. To address this issue, we therefore turn to the analysis of the effect technical interdependencies within each module.

5.6.2. Within-Module Analysis – Panel Regressions

Table 4 reports the estimates from the panel regression analysis. The standard errors are clustered by module to account for non-independence of observations within each module.

Model 1 (second column from the left) is highlighted in bold as it serves as the foundation of our theorization below. The main and control effects in the model are consistent with our expectations. The coefficients of the main effects of *design patterns* are the population average and denote the intercept of all modules in that design pattern⁴⁰.

The effect of *external dependencies* is positive and significant ($\beta = 0.028$, $p < 0.01$). This shows that increasing the number of external dependencies has - on average across all panels - a negative effect on the pace of change as it increases the time intervals between changes to functionality. This is very much in line with modular systems theory that suggests the coupling of modules has detrimental effects for the performance of a system. The control for technical integration (i.e. lines of code per source code file) decreases the time intervals between functionality changes ($\beta = -0.015$, $p < 0.01$). This indicates that the pace of changes increases as source code files grow in relative size hence modules are becoming more integrated. The second control for *technical debt* has no effect ($\beta = 0.002$, $p > 0.1$) beyond the controls-only baseline model (left-hand column). This indicates that changes in module size relative to the number of evoked interdependencies is inconsequential for the pace with which module functionality changes. The third control for *secondary changes* has a positive

⁴⁰ The lack of significance of the intercepts is inconsequential to our theorising

and significant effect ($\beta = 0.124$, $p < 0.01$). This result highlights the difference between the kinds of changes affecting a module. The control assures us that the time intervals between functionality changes are meaningful as, in general, the time between functionality changes increases as the result of second-degree development work such as maintenance, documentation, or refactoring.

Lastly, the control for *change rate* presents a negative significant coefficient ($\beta = -0.335$, $p < 0.01$) indicating that the overall rate with which a module is developed affects the speed of its evolution. This in itself is not surprising, but since the effect is significant on the population level it indicates that all modules are subject to that effect. In other words, the effect of the overall change rate does not explain the relative differences we observed across design patterns in the first step of the analysis.

The main intention of modelling the effect of external dependencies is to capture the role of external dependencies for modules in each of the design patterns that contrast classical modularity in our framework. Here, the effects of the interactions between design patterns and number of external dependencies reveal interesting findings.

Hypothesis H1 stated that modules whose designs deviate from established modular design principles have a slower pace of change. We find support for this hypothesis. There is a significant effect of increasing the number of external dependencies in modules with *high structural complexity* and *low external dependence* ($\beta = -0.019$, $p < 0.01$). Yet, despite observing an effect in these modules, the comparison across modules implies that the pace of change remains slowest relative to modules in all other design patterns. This finding aligns with expectations in modular systems theory as it indicates that unboundedness does not remedy flaws in module design such as an overly complex architectural structure. As a result, although these modules benefit from externalizing some of their dependencies to source in resources from the ecosystem, their overall performance does not rival that of other modules.

In contrast, hypothesis H2a is rejected. We assumed an increase in external dependence for modules would uniformly benefit the pace of change in modules regardless of structural complexity. However, as captured by the effect of increasing external dependencies for modules with the design pattern *low structural complexity and high external dependence* effect here is not significant ($\beta = -0.01$, $p > 0.1$). This demonstrates that modules with low structural complexity do not benefit from adding

external dependencies. In addition, the overall performance of these modules is mediocre relative to all other modules. This suggests that externalizing dependencies to sources in the ecosystem does not foster a module's ability to change if modules comprise non-complex designs to begin with. This implies that enforcing unboundedness for module designs that could potentially provide the intended functionality within their boundaries is not beneficial.

Lastly, we find strong support for hypothesis H2b. The coefficient of the interaction term between the design pattern of *high structural complexity and high external dependence* is negative and statistically significant ($\beta = -0.027$, $p < 0.01$). Hence, the greater the number of unique external dependencies used in structurally complex platform modules, the shorter the time intervals between changes in their functionality. Also, when compared to other designs, modules with *high structural complexity and high external dependence* present a pace of change that is significantly higher. Modules in these design patterns derive a large part of their structural complexity from external dependencies. The finding suggests a deviation from classical modularity as module architecture evokes substantial couplings both within and across modules.

Since we estimated unstandardized coefficients, the findings can be expressed in real unit terms. Assume a module in the design pattern *high structural complexity - high external dependence* is associated with an average change time of 10080 minutes in t_1 (i.e. one function change per week). In this hypothetical example, an external dependency added to the code base shortens the time interval between functionality changes by 1620 minutes (~27 hours), the equivalent of a 16.1% increase in the pace of change over t_1 ⁴¹. Figure 5.3 below illustrates our findings that will be discussed in the last section.

Figure 5.3. Model Results in 2x2 Framework

⁴¹ Since $\exp(\log(10080) * (1-0.019)) = 8460$ (i.e. a difference of 1620 minutes or 16.1% of 10080)

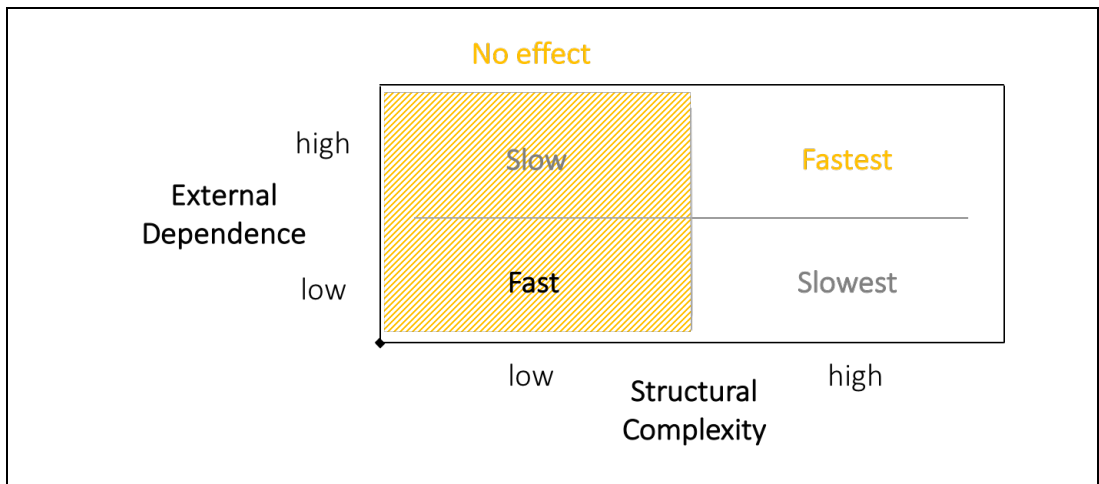


Table 5.4. Model Results – Mixed Effects Panel Regressions

Predictors	Dependent variable									
	Change Time (all functions) <i>Baseline</i>		Change Time (all functions) <i>Model 1</i>		Change Time (new functions) <i>Model 2</i>		Change Time (all functions) <i>Model 3</i>		Change Time (new functions) <i>Model 4</i>	
Design: High - High			0.299	(0.680)	0.481	(0.730)	0.435	(0.672)	0.737	(0.712)
Design: High - Low			0.487	(0.696)	0.507	(0.732)	0.809	(0.669)	0.817	(0.694)
Design: Low - High			-1.669**	(0.732)	-1.955**	(0.768)	-1.421*	(0.710)	-1.714**	(0.734)
External Dependencies			0.028***	(0.006)	0.028***	(0.006)				
External Dependencies (strict)							0.019***	(0.004)	0.019***	(0.004)
Technical Integration	-0.021***	(0.004)	-0.015***	(0.004)	-0.016***	(0.004)	-0.014***	(0.004)	-0.014***	(0.004)
Technical Debt	0.006**	(0.002)	0.002	(0.003)	0.002	(0.003)	0.003	(0.002)	0.003	(0.002)
Secondary Changes	0.130***	(0.010)	0.124***	(0.009)	0.118***	(0.010)	0.125***	(0.009)	0.119***	(0.010)
Change Rate	-0.328**	(0.049)	-0.335***	(0.049)	-0.352***	(0.052)	-0.334***	(0.048)	-0.353***	(0.051)
Design: High - High x External Dependencies			-0.019***	(0.007)	-0.023***	(0.007)				
Design: High - Low x External Dependencies			-0.027***	(0.006)	-0.028***	(0.006)				
Design: Low - High x External Dependencies			-0.01	(0.006)	-0.008	(0.006)				
Design: High - High x External Dependencies (strict)							-0.012***	(0.005)	-0.015***	(0.005)
Design: High - Low x External Dependencies (strict)							-0.019***	(0.004)	-0.018***	(0.004)
Design: Low - High x External Dependencies (strict)							-0.004	(0.004)	-0.002	(0.004)
Constant	12.450***	(0.244)	11.578***	(0.508)	11.718***	(0.532)	11.188***	(0.512)	11.344***	(0.530)
Observations	1,388		1,388		1,260		1,388		1,260	
Groups	16		16		14		16		14	
Marginal R-Squared (%)	17.8		29.9		29.2		29.2		28.3	
Conditional R-Squared (%)	35.7		52.3		53.1		49.0		49.2	
Note:	*p<0.1; **p<0.05; ***p<0.01									
	<i>robust standard errors in parentheses (clustered by module); first order autoregressive covariance correction for all models</i>									

5.6.3. Robustness Checks

We implemented a number of robustness checks to increase confidence in the reported effects. First, we controlled for random effects over time by estimating the variance of a random slope component. This yielded a value very close to zero indicating a negligible effect per module (e.g., in Model 1: $\sigma^2 \approx 7.11\text{e-}12$).

We repeated the analysis with an alternative dependent variable that only measured time intervals between changes that add new functionality [DV: *Change Time (new function)*] and achieved consistent results (*Models 2, and 4*). Although *Model 2* exhibits a slightly higher proportion of variance explained, we base our theorization on *Model 1* for two reasons. First, the dependent variable in *Model 1* is more encompassing as it captures all changes made to module functionality instead of only functions that are newly added. Second, the change in the dependent variable in *Models 2(4)* leads to the exclusion of two modules from the sample with insufficient amount of observations.

Additionally, we used an alternative measure of external dependencies that strictly treated every dependency in the source code as external if it was not part of the respective module and hence implies it has not been created by the module developer. This classifies dependencies as external even if they are standard libraries in a programming language or included in ‘off-the-shelf’ development software. Despite this severe adjustment, all effects remained consistent throughout albeit slight reductions in effect size (*Models 3, and 4*)⁴².

Finally, we ruled out two sources of confounding bias unique to the empirical context. On the one hand, we confirmed that the used programming languages are represented across all design patterns. For instance, modules written in the popular programming language *Java* are found in each of the design patterns. This indicates that individual differences in the used programming language are an unlikely source of unobserved bias in the formation and behaviour of the design patterns. An important insight, since all of the used programming languages are instances of the object oriented paradigm and thus share similarities in computational and representational logics (Gamma *et al.*,

⁴² The used classification of the source of dependencies in models 1 and 2 is far more realistic than the ‘strict’ classification that models 3 and 4 are based on.

1994). We also checked that modules are evenly spread across design patterns with respect to their function. Using descriptions of main functionalities, we made sure software such as editors or data handling tools are not lumped together in one design pattern. A complete overview can be found in the appendix.

5.7. DISCUSSION

The paper set out to reflect on modularity as the basis for understanding change in digital platform ecosystems. In particular, we argue that digital platform modules exhibit a number of qualities – agnosticism, programmability, and interactivity – that differ from the physical components typically studied in modular product systems. As a result, digital platform ecosystems differ from modular systems in that they are characterized by unboundedness. The propensity to stimulate novel and often unanticipated interactions across architectural boundaries (Eaton *et al.*, 2015; Parker *et al.*, 2017) are salient indicators for the increasing unboundedness of platform ecosystems.

However, current studies do not pay enough attention to the importance of unboundedness of digital platform ecosystems. Whether it is the smartphone or the solar industry, modules are thought of as something “that together comprise a coherent solution” (Hannah and Eisenhardt, 2018). This is consistent with recent theorizing on ecosystems (Jacobides *et al.*, 2018), in which modularity, largely understood as design hierarchies (Clark, 1985) is seen as a backbone. Yet, we know that many digital technology modules transcend the original focal product system, not only as stand-alone products, but also as components across inherently different platform ecosystems (e.g., Henfridsson *et al.*, 2018)

The traditional view of modularity suggests that a high pace of change requires loose coupling between modules. With the idea of design hierarchies in mind, simplicity in the interfaces between modules is central when the platform operator is assumed to directly govern the design rules (Baldwin and Clark, 2000). If there were tight couplings, the argument goes, prohibitively high coordination cost would hamper the ability to change. Contrary to this traditional view of modularity, we propose and test a set of hypotheses suggesting that unbounded platform ecosystems work differently, and defy the notion of structural complexity and external dependence as unequivocal

disadvantage to a system's performance. It is thus important to sort out the consequences for how to think about complexity and its relation a digital platform's ability to change.

With this reflection on modularity of digital technology, our study contributes theoretically and empirically to the literature on platform-based ecosystems (Jacobides *et al.*, 2018; Kapoor and Agarwal, 2017; Wareham *et al.*, 2014). Modularity is one of the most important intellectual underpinnings of platform ecosystems (Jacobides *et al.*, 2018; Gawer, 2014; Tiwana *et al.*, 2010). Modularity's core promise is to increase the ability of a technical system to change by hiding complexity. Indeed, modularization across a platform's ecosystem is a key driver for success as it increases the ability to incorporate change to the platform (Tiwana, 2015). Yet, in order to foster change, platform operators are advised to adhere to modular design principles, first and foremost demanding that dependencies between modules are best kept to a minimum. This suggestion underemphasizes what makes digital technologies unique; modules comprising software are system-agnostic, reprogrammable, and interactively-designed. As a consequence, digital platform ecosystems become unbounded systems and require little to no coordination to resolve interdependencies. Instead these systems rely on technical interoperability as the premier condition for interaction (Jacobides *et al.*, 2018, McIntyre and Srinivasan, 2017).

This unboundedness of ecosystems has implications for organisations engaged in technology platforms. The objective of digital platform ecosystems is to promote complex interactions so as to stimulate serendipitous interactions made possible by the attributes of the underlying modules (Wareham *et al.*, 2014; Yoo *et al.*, 2010). As such, our findings frame a strategic option for platform-operators. While resolving module interdependencies remains an important and valid course of action (Ethiraj and Levinthal, 2004), our analysis reveals an additional lever for platform strategy. Platform operators can manage unboundedness to stimulate the pace of change in core modules. As our empirical analysis shows, modules that embrace unboundedness in digital platform ecosystems - by drawing from external libraries or appropriating framework extensions - increase the pace of change in modules despite an increasing structural complexity. The unbounded character of digital technologies does not constrain architectural designs to the same extent as classical modular systems. For one, the features encapsulated by external dependencies makes changes to module

functionality easier to implement even if that induces structural complexity. Furthermore, building external dependencies in modules aligns platform activity with technologies outside the platform, reaping external performance improvements while minimizing development effort for the focal module (Tiwana, forthcoming).

This suggests that platform operators gain performance advantages by promoting unboundedness across the platform as core modules externalize structural complexity. Participants in digital platform ecosystems often leverage the shared architectures of digital technologies to draw from capabilities developed elsewhere. The platform thus benefits as swift innovations are enabled through increased interaction in shared architectures (McIntyre and Srinivasan, 2017). Here, our study complements earlier research on innovation activity in platform ecosystems given an increase in openness and interoperability (e.g., Boudreau, 2010). It also aligns the study with recent work demonstrating that structural complexity and performance are by no means mutually exclusive if interaction is driven by shared artifacts across the platform's ecosystem (e.g., Tiwana, forthcoming; Kapoor and Agarwal, 2017).

Ultimately, our findings suggest a reflection of the view on digital technology in research on platform ecosystems. Beyond deterministic views on managing technology and innovation, scholars rarely recognize that digital technology is different not only in degree but also different in kind. Our conceptualization of modules in digital ecosystems addresses the need for a deep understanding of unique aspects of digital technology and the implications arising from organizing with digital technology. Established principles in modularity may have to be carefully revisited to deepen our understanding of the full implications of a digitalized world. Our study offers a starting point for organisation scholars to take digital technology more seriously.

APPENDIX TO CHAPTER 5

Technical note on source code queries

We used the regular expression engine as implemented in the statistical programming language *R* (Version 3.3.2)⁴³ for text queries from the module source code. An overview of the main query terms and allowances for variation in the code are provided in the table A1 below. For function calls we focus on named functions that are available for reuse if developers decide to use the module for their work. Functions in code come in a large variety. The approach reported here excludes functions used in ad-hoc computations (‘anonymous functions’). We relied on complementary data sources to guide the analysis. For instance, to decide on which queries to include during the text analysis, we used technical documentation pertaining to the represented programming languages. As such, we queried for code expressions that were in line with conventions of the respective programming languages. We also consolidated our source code text queries with expert users of the respective programming languages. Included in the below table is an overview of typical dependencies found in the empirical context and how the study classified them as either internal or external to the respective platform module.

⁴³ Documentation available at <https://cran.r-project.org>

Table 5.5.A. Module Overview

Module Name	Programming Language	Design Pattern	Main Task	Internal Dependencies	External Dependencies
mod_tile	C	low-low	data export	standard C libraries loaded in header files such as: math.h; stdio.h; string.h, or time.h	non-standard libraries: postgresql/mysql; mapnik renderer; boost library
osm_website	Ruby	low-low	data presentation	standard libraries (gems) such as: base64, date, expect, json	non-standard libraries: libxml; pagination: oauth / openid: mysql; jquery
osmosis	Java	low-low	data transform	standard java libraries such as: java.lang; java.util; java.io; java.text; java.math	non-standard libraries and extensions such as: gradle; xml.sax; postgis; postgresql; java.awt
planetdump	C	low-low	data export	standard C libraries loaded in header files such as: math.h; stdio.h; string.h, or time.h	pqxx; mysql
potlatch	ActionScript	low-low	data edit	built in packages such as; SWT	SWT library; ming 0.3
cgimap	C++	high-high	data export	standard C/C++libraries loaded in header files such as: cassert.hpp; cfloat.hpp; cstdlib.hpp	non-standard libraries: libxml2; libpqxx; libfcgi; libboost; libcrypto++; mysql/ssqls
gosmore	C++	high-high	data presentation	standard C/C++libraries loaded in header files such as: cassert.hpp; cfloat.hpp; cstdlib.hpp	xmlwriter; android-log; GLES package
osm2pgsql	C++	high-high	data transform	standard C/C++libraries loaded in header files such as: cassert.hpp; cfloat.hpp; cstdlib.hpp	non-standard libraries: proj.h; boost lib; libpq-fe; osmium, geos, protozero
planet_gpx_dump	Python	high-high	data export	standard python libraries: e.g. document; os; sys; datetime	psycopg2; libxml

splitter	Java	high-high	data transform	standard java libraries such as: java.lang; java.util; java.io; java.text; java.math	extensions such as: libxml; java.awt
iD	JavaScript	high-low	data edit	built in javascript objects such as: String, Math, Array, JSON	extensions such as: d3; xml2js
josm	Java	high-low	data edit	standard java libraries such as: java.lang; java.util; java.io; java.text; java.math	java.awt/swt; oauth; google- signpost; kitfox/svg; apache validations; gnu/getopt
Nominatim	C	high-low	data presentation	standard C libraries loaded in header files such as: math.h; stdio.h; string.h, or time.h	libpq-fe; boost lib; pear-db; sqlite; mysql
merkaartor	C++	low-high	data edit	standard C/C++libraries loaded in header files such as: cassert.hpp; cfloat.hpp; cstdlib.hpp	qt framework libraries; libraries such as: ggl xmllib; zip; google auth; gdal; proj
mkgmap	Java	low-high	data export	standard java libraries such as: java.lang; java.util; java.io; java.text; java.math	parabola lib
potlatch2	ActionScript	low-high	data edit	built in packages such as; flash.xxx	flex libraries; oauth; fcs;h; adobe/as3 extensions

Table 5.6.A. Queries used in Source Code Analysis

Programming Language	Query terms for functions	Query terms for dependencies	Variations in query
ActionScript	function (abc) { xyz }	import xyz	space and/or line breaks before query term and between brackets
C	return_type function_name (abc){xyz}	#include xyz #import xyz	space and/or line breaks between brackets; adjustments for implicit/explicit inline function definitions
C++	return_type function_name (abc){xyz}	#include xyz; #import xyz	space and/or line breaks between brackets; adjustments for implicit/explicit inline function definitions
Java	return_type function_name (abc){xyz}	#import xyz	space and/or line breaks between brackets
JavaScript	function function_name (abc) {xyz}	import xyz; require(xyz);	Node.js syntax considered
Perl	sub {xyz}	use xyz; require xy	space and/or line breaks before query term and between brackets
PHP	function function_name (abc) {xyz}	include xyz; include_once xyz; require xyz; require_once xyz	space and/or line breaks before query term and between brackets
Python	Def function_name (abc):	import xyz	NA
Ruby on Rails	def abc	require xyz; include xyz; require_relative xyz	NA

CHAPTER 6 – PERICENTRIC COORDINATION ON DIGITAL PLATFORMS DURING BOUNDARY RESOURCE INSTABILITY

ABSTRACT

Digital platforms often require contributions from platform participants. Thanks to stable and standardised interface capabilities, such as boundary resources, participant can contribute without excessive coordination with the platform operator. Using a computationally-intensive mixed-methods approach, we study the geo-data platform OpenStreetMap. We focus on two episodes of boundary resource instability i.e. situations in which the basis for how interaction is facilitated between a platform and its participants changes drastically and is temporarily unpredictable. We explore how participants coordinate their contributions in such situations and find a mode of coordination that unfolds whenever participants require additional information in order to proceed with their contributions. Unaccounted for in the literature, we theorise this mode of coordination on digital platforms as “pericentric coordination”. Pericentric coordination is observable due to instability in the interfaces that would normally facilitate interaction between participants and the platform. In such situations, many otherwise disengaged participants are forced to engage in coordination before they can return to their contributions. We characterise this mode of coordination by identifying four salient attributes as; (i) inclusive beyond a mere core-periphery distinction, (ii) occurring sporadically over time, (iii) benefitting participant contributions, and (iv), complementary to other forms of coordination.

The pre-study to this chapter was presented as

Hukal, P. (2018). Network Structure and Digital Platform Development: Showcasing the Generation of Research Questions through Inductive Analysis of Trace Data

Academy of Management Specialized Conference: Big Data and Managing in a Digital Economy, 18-20 April 2018, Surrey, UK

The pre-study is enclosed as an addendum to this chapter.

PERICENTRIC COORDINATION ON DIGITAL PLATFORMS DURING BOUNDARY RESOURCE INSTABILITY

A pericenter is the point at which any orbit is nearest to the center of its attraction
- *From the Glossary of the*
American Meteorological Society

6.1. INTRODUCTION

Many digital platforms rely on the contributions of distributed and heterogeneous participants. Participant contributions come in various forms including developing applications (e.g., iOS App Store), providing content (e.g., YouTube), or sharing data (e.g., OpenStreetMap). The success of digital platforms thereby rests on the limited coordination effort that is necessary to elicit such participant contributions. The scale on which many digital platforms operate would make coordinating individual contributions prohibitively cumbersome.

In order to facilitate participant contributions, digital platforms rely on stable and standardised interfaces on the platform boundary (Eaton *et al.*, 2015; Ghazawneh, 2012; Ghazawneh and Henfridsson, 2013). What is referred to as boundary resources (Eaton *et al.*, 2015; Ghazawneh and Henfridsson, 2013), includes artifacts such as application programming interfaces (APIs), software development kits (SDKs), or licensing agreements. Boundary resources are hence the capabilities that serve as the interface between a platform and its participants. As they facilitate interaction between the platform and its participants, boundary resources are key components in enabling participant contributions (Eaton *et al.*, 2015; Ghazawneh and Henfridsson, 2013; Karhu *et al.*, 2018).

In the same time, an important objective of boundary resources is to minimise the effort needed to coordinate participant contributions. On digital platforms, a large part of coordination activity is abstracted to the design of interfaces on the platform boundary. A typical platform boundary resource, such as an API used by third-party developers, facilitates contributions from platform participants. Coordination is thereby pre-empted as the intentions of the platform operator are already reflected in functionality and design of the interface (Ghazawneh and Henfridsson, 2013).

In order for participants to continue to contribute with minimal coordination, boundary resources need to be stable. Boundary resource stability means that the logic by which interaction between platform and participants remains unchanged so that recurrent

interaction with the platform is possible. Stability of boundary resources translates into reliability for platform participants as their engagement with the platform is seamless and uninterrupted by coordination. As such, boundary resources stability is crucial for the smooth functioning of a platform and platform operators have an immense interest in the stability of their deployed boundary resources (Eaton *et al.*, 2018; Karhu *et al.*, 2018). In addition, stability of a boundary resource supports platform growth and innovation as a function of the continuous interaction it enables with the platform. Stability is crucial as invariant resource availability forms the basis for recombination and so eventually favours the creation of novel derivatives on the platform (Yoo *et al.*, 2012).

It is therefore important to realise that stability is one determinative factor for how conducive a boundary resource is and how much coordination can regularly be substituted by such interfaces. In this regard, it is imperative to further investigate the relationship between coordination on digital platforms and the stability of boundary resources. In absence of stability, boundary resources cannot absorb coordination efforts. By extension, this means unstable boundary resources are less reliable, making interaction with the platform temporally unpredictable for platform participants. We therefore ask the following research question:

How do participants on digital platforms coordinate contributions during boundary resource instability?

We address this research question in the context of the geo-data platform OpenStreetMap. Using the OpenStreetMap platform as an illustration, we report an exploratory case study drawing from a computationally-intensive mixed-methods approach (Berente *et al.*, 2018; Venkatesh *et al.*, 2016). We identify two situations of boundary resource instability; (i) the introduction of a new RESTful web API, as well as (ii) the change of the contributor license that regulates how participants can use and re-use geo-data. As important boundary resources facilitating participant contributions, we explore participant coordination and the contributions on the platform during times when both boundary resources are subject to changes and hence unstable. That is, in situations in which the basis for how interaction is facilitated

between a platform and its participants changes drastically and the logic of how the boundary resource facilitates interaction is altered unpredictably.

We collected a rich data set of communication and development logs of individual platform participants. Our findings indicate a mode of coordination hitherto untheorized in the platform literature; “pericentric coordination”. Pericentric coordination denotes activity during meaningful yet rare touchpoints that enable participants to return to their contribution work. Such coordination is relevant when platform participants can no longer contribute without friction. In these situations, participants engage in coordination to chart out a way forward in order to continue to contribute to the platform’s operation.

This study thus makes an important contribution to the literature on digital platforms. We attend to participant dynamics during times of boundary resource instability and provide insight into a form of coordination that has been overlooked thus far. We thereby add a description of participant coordination that extends core-periphery distinctions as expected in adjacent literature. Instead, we highlight how platform participants engage when situations of instability in the boundary resources necessitate coordination before returning to their individual contributions. Platforms normally can do without such coordination but require well defined and stable interfaces to do so. We find that in times of boundary resource instability, a group of participants is active that engages little in coordination work, but contributes heavily throughout. We are thus able to deepen our understanding of participant dynamics shaping growth on digital platforms.

The paper proceeds with a review of related work platform boundary resources and derives expectations towards coordination on digital platforms. We then introduce the empirical setting of the case study and highlight results from our exploratory analysis. We finally theorise our findings by formulating four propositions about the coordination among platform participants when platform boundary resources are unstable.

6.2. CONCEPTUAL BACKGROUND

6.2.1. Platform Boundary Resources

Boundary resources have been highlighted as important strategic components capable of guiding participants and their contributions on a platform (Eaton *et al.*, 2015; Ghazawneh and Henfridsson, 2013; Karhu *et al.*, 2018). While the interaction between platform and participants is generally guided by interfaces describing how components connect and communicate (Baldwin and Clark, 1997, 2000; Garud and Kumaraswamy, 2003), boundary resources enable such connection and communication across the platform boundary (Ghazawneh and Henfridsson, 2013). For instance, a RESTful web API facilitates information flows between platform, platform participants, and other artifacts in such a way that participants can draw on platform resources for the creation of novel services while ensuring future interaction across the platform (Boudreau, 2017).

Initial work on boundary resources has highlighted how acts of “resourcing and securing” the platform contribute to its growth (Ghazawneh and Henfridsson, 2013). The key task of boundary resources is therefore the continuously-iterative regulation of which and how resources are available to platform participants. Expanding platforms by offering a resource followed by the subsequent securing by establishing rules for interaction thus manages platform growth through boundary resources (Ghazawneh and Henfridsson, 2013).

Boundary resources thereby present nexuses of interaction between platform and platform participants. As a consequence, such interfaces evolve during a platform’s lifespan. For instance, Eaton *et al.* (2015) explicate how interaction between heterogeneous, distributed participants and platform operator shape boundary resources over time. In reaction to tussles arising from initial unilateral designs, they observe incremental changes in boundary resources design and functionality as diverse participant interests are successively reflected. They describe this change as a tuning process wherein boundary resources are affected by “cascading actions of accommodations and rejections of a network of heterogeneous actors and artifacts” (ibid:p. 217).

Recently, Karhu *et al.* (2018) suggested to extend the understanding of boundary resources by examining the kind of interaction such resources provide. In their study,

they distinguish boundary resources by access to existing resources and the creation of shared resources. The distinction offered by Karhu and colleagues thereby speaks to consequences of different forms of openness presented on platforms and how such forms are orchestrated through the use of boundary resources (Karhu *et al.*, 2018).

Crucial to the efficacy of boundary resources is their stability. Stable interfaces offer some oversight to the platform provider while utilising a platform's potential for interaction with participants (Baldwin and Woodard, 2009; Tiwana *et al.*, 2010). In order for participants on a platform to engage without the operators involvement, the platform needs to exhibit some degree of inertia (Wareham *et al.*, 2014). The kind and extent of interactions that the platform invites, need to be stable to a degree that enables actors to readily and repeatedly engage with the platform (Ghazawneh and Henfridsson, 2013; Tilson *et al.*, 2010). Given that boundary resources mediate access to platform resources, the need to do so reliably is crucial for platform growth. The promise of recurrent and unchanged interactions seems important as reliability is the basis on which a platform grows and innovates. For instance, stability of a boundary resource contributes to platform growth. That is, using and reusing platform resources is a function of recurrent interactions enabled through boundary resources. Stability in such interfaces is therefore crucial for growth and innovation on platforms as invariant resource availability forms the basis for recombination and so eventually favours the creation of novel derivatives (Yoo *et al.*, 2012).

6.2.2. Participant Coordination on Digital Platforms

With the aim to reliably facilitate interaction, boundary resources are important means of participant coordination on digital platforms. We follow Malone and Crowston and refer to coordination as “*additional information processing performed when multiple, connected actors pursue goals that a single actor pursuing the same goals would not perform.*” (Crowston, 1997; Malone, 1988).

Coordination theory informs a large part of research in the domain of digital technology mediated work (Crowston, Rubleske, and Howison, 2004). Pertinent for our investigation is the notion that distributed, independent, and heterogeneous participants contribute to platform operations by engaging with the work of others. This dynamic is well studied in the literature on the open source software development phenomenon (Crowston, 1997; Crowston *et al.*, 2004; Lindberg *et al.*, 2016). While

this is adjacent to most digital platform studies, this body of work can meaningfully inform inquiries into participant contributions on digital platforms.

In this stream of literature, several mechanisms have been described by which participants coordinate their work in an arm's length manner. The fact that coordination is rendered by digital technology is thereby non-trivial. Baldwin and Clark (2006), argue that the participation and contribution of heterogeneous, temporally as well as spatially distributed developers is a direct consequence of the digital technology and its representation: source code. In conjunction with the modularization of source code, it is the digital representation of information in text form that allows participants to access, re-use, and alter software code with little need to engage in upfront coordination (Baldwin and Clark, 2006).

The modularisation of code has consequences for the interaction it enables among participants wishing to contribute to ongoing developments. One example in past work has demonstrated how individual contributions defer work on seemingly overwhelming problems in the codebase until the cumulative effort of other participants enables addressing issues of the earlier complex problem (Howison and Crowston, 2014). Referred to as "open superposition", this dynamic allows for most codebases to advance through piecemeal contributions in the absence of coordination through hierarchies, teams, or markets (Howison and Crowston, 2014).

Another example illustrates how the use of digital tools and the engagement in development work itself is the mechanism that substitutes coordination. In their research on legitimate peripheral participation, Gasson and Purcell (2018), illustrate how digital tools deployed in a distributed development context surrogate coordination tasks. Coordination herein unfolds by participants showcasing their skill with the tools and practices instead of direct communication (Gasson and Purcell, 2018). Others have picked up that theme as well. Shaikh and Vaast (2016), describe the processes of development coordination and realisation as an oscillating between carrying out tasks in the public sphere of a project in contrast to getting work done in more closed-off, private arrangements using idiosyncratic tools and procedures (Shaikh and Vaast, 2016).

However, some problems persist even if digital technology mitigates issues in the coordination of work. For instance, Lindberg *et al.* (2016) observe increased complexity and diversity of activities among developers. They theorise that changes in the interdependent parts of a digital technology requires increased coordination as interdependencies have to be resolved or mitigated. In particular, they refer to these dynamics as a form of “collective sense making, where a diverse set of developers jointly seek to achieve common ground” (Lindberg *et al.*, 2016; p. 764).

Coordination in most of these studies is assumed to unfold through the division of participants in a core and a periphery group. Indeed, a core-periphery distinction in open source projects is well-established (Crowston and Howison, 2006). Scholarly attention has mainly rested on the conjecture that participants become part of a core group based on individual contributions. The rationale is that learning effects result in the specialization of developers simply as a function of the number of tasks undertaken within a problem domain (Crowston and Howison, 2006). This continued demonstration of technical expertise legitimizes the work of individuals over time. Beyond the amount of effort put into a project, the nature of individual contributions further qualifies participants as belonging to one or the other group. For instance, the engagement with other developers and the position in the social structure of a participant group is deemed determinative of core membership (Dahlander and Frederiksen, 2012).

As participants contribute to the technical development of a project, they signal their knowledge and competence. This furthers their progression to what is considered core, and, more crucially, means participants take over more and more coordination tasks (Dahlander and O’Mahony, 2011). A salient expectation is thus the existence of an equivalency between coordination and development work. That is, participants that are considered to be core, contribute to both the coordination of the work as well as the work itself.

Not only do open source software projects vary widely across all conceivable attributes such as group size of participants, participant organisation, social structure of participants. Perhaps more importantly, digital platforms follow dynamics in participant coordination that differ from open source projects. To the extent that platform boundary resources work in a stable and foreseeable manner, platform

operators have a powerful tool at their disposal that coordinates participant contributions without direct engagement of the operator. At the very least, boundary resources maintain control for the platform operator by regulating access to the platform (Ghazawneh and Henfridsson, 2013). This limits coordination as possible pathways for participants to engage with the platform are predefined by the operator. Therefore, boundary resource moderate and facilitate the interaction of participants and artifacts across the platform without direct involvement of the platform operator. This pre-empts coordination activities as outcomes of information exchanges are already embedded in make-up and design of boundary resources – even if such attributes are being renegotiated down the line (e.g., Eaton *et al.*, 2015). Whether as components or framework agreements, access to and creation of participant contributions is thus guided by the boundary resources on platforms (Karhu *et al.*, 2018).

However, with the mission of better understanding participant coordination on digital platforms, further work is needed into situations in which boundary resources are not stable. On digital platforms, coordination is abstracted to the design and functionality of resources on the platform boundary. Yet, we know little about situations in which platform participants can no longer rely on the interfaces provided by digital platforms that enable participants to contribute. In these situations, participants would be forced to engage in coordination work as the boundary resource is no longer providing the reliability that makes a lot of coordination work on digital platforms superfluous.

Paying attention to the emergent dynamics in coordination on platforms aligns with past work using coordination theory to understand collaborative work mediated by complex networks of digital technologies (Lindberg *et al.*, 2016; Lyytinen *et al.*, 2016). In particular, the study contributes to the understanding of the dynamics of participant coordination on digital platforms that unfold when boundary resources are unstable (Eaton *et al.*, 2015; Karhu *et al.*, 2018).

6.3. RESEARCH DESIGN

6.3.1. Empirical Setting

Our study investigates participant coordination of work during times of boundary resource instability in the context of the OpenStreetMap geo-data platform. Referred

to as the ‘Wikipedia of maps’ (Fox, 2012), the platform provides geo-spatial data at a global scale online and for free. In an informant interview, a participant who is engaged in development work on the platform referred to the OpenStreetMap as essentially being “*a massive spreadsheet holding geo-location data*”. Typically, participants rely on a variety of tools in order to edit, retrieve, format, and store geo-data (Ramm *et al.*, 2010). While participants interact with the platform in a variety of ways, three modes are salient:

- (i) Mapping: providing, editing, storing geo-data on the platform with the aim to provide detailed mapping data,
- (ii) Developing: creating and maintaining the software tools available to work with geo-data,
- (iii) Using: drawing from the OpenStreetMap geo-data repository to power diverse web services.

The context is well suited to study the coordination of work among participants on digital platforms. All of the above use cases are enabled through stable boundary resource interfaces on the platform. *Mapping* draws from a standardised data model that prescribes the meta-structure through which geo-data is being stored and retrieved (see e.g., Ramm, 2015). *Developing* leverages interface capabilities between software components on the platform, for instance, in the form of APIs or standardised protocols (Ramm *et al.*, 2010). *Using* relies on the access to geo-data which enables the use and re-use of contributed geo-data for the creation of commercial and non-commercial web services, for example, in road navigation applications.

6.3.2. Data

We collected data from two sources with major relevance to the coordination of work in the context of OpenStreetMap.

First, we collected data from the archive of the mailing list used by participants. The mailing list is one of the core points of contact⁴⁴ used by mappers, developers, and application service providers to ask questions, discuss changes, or announce recent developments. The mailing lists is divided into thematic sub-lists. To focus the

⁴⁴ https://wiki.openstreetmap.org/wiki/Mailing_lists; last accessed 25 May 2018

analysis, we concentrated on the mailing sub-lists pertaining to technical development work. Using a custom-built script, we downloaded all e-mail messages including related metadata from 19 topic lists with a technical subject matter (we provide a full list and descriptions in the appendix). The entire dataset spans the time from 2005 to 2018, and captures communication among ~1,000 participants, in 15,000 e-mail discussions consisting of ~72,000 separate messages. The e-mail messages are grouped by subject and thus present discussions in a separated, threaded structure.

Second, we collected source code development data from the platform's GitHub repositories. Participants on OpenStreetMap are encouraged to contribute to the technical developments on GitHub where the platform's repositories hold source code needed to operate the platform as well as add-on software tools. GitHub is based on the Git version control system and lets users share, propose, and discuss software code changes (Dabbish *et al.*, 2012; Tsay *et al.*, 2014). Herein, platform participants interact on the basis of collaborative source code production (Baldwin and Clark, 2006). Specifically, the data we collected are so-called "commits". On GitHub, users can suggest source code changes. If a suggested source code change accepted, it is "committed", and thus merged with the original codebase and the alteration takes effect. GitHub makes the content of every commit as well as metadata publicly available. Using the GitHub web API, we downloaded all source code changes (~59,000) committed to OpenStreetMap software components (n=19) between December 2006 and June 2017.

Our main interest is the interaction between participant coordination work as captured in e-mail discussions and their contribution to the technical development of the platform as reflected in their commits to the code base. We aim at understanding how participants on the platform coordinate their contributions during times when the stability of boundary resources is at risk. As such, our unit of analysis is the individual participant who is engaged in both, communication and development work. Between the data collected from the two sources described above, we capture coordination efforts through participant activity in communication and development. Upon identification of developers in both datasets, we are able to relate communication data to observations of how participants contributed to the development work on the platform.

We complement the two data sources above with insights from informant interviews and secondary data from interviews in mainstream publications. Informant interviews consisted of unstructured interviews with a range of platform participants. The role of participants varied from mappers, developers, to external service providers and were conducted to complement archival evidence with rich contextual information. Table 6.1 below gives an overview over the data sources used in the analysis. We detail informant roles and interview topics in table 6.5 in the appendix.

Table 6.1. Data Collection			
Data Source	N	Description	Utilisation
E-mail archive	~72,000	Archived e-mail conversations among platform participants (mappers, developers, external users) with a focus on topics of technical development	Information exchanges in the e-mail archive capture coordination work among participants on the platform
Software development data	~59,000	Development data (commits); a commit is the smallest unit of technical alteration on GitHub as source code changes are incrementally <i>committed</i> to the code base	Technical development of software used to operate the platform reflect individual contributions by platform participants
Informant interviews (primary)	9	Unstructured interviews with platform participants; see table A 6.2 in the appendix	Interviews complement and contextualise interaction among of platform participants
Informant interviews (secondary)	15	Interviews with participants published in an edited volume: Coast (2015)	

6.3.3. Analysis and Framing

We adopted a computationally-intensive inductive approach (Berente *et al.*, 2018) following a sequential mixed-methods design (Venkatesh *et al.*, 2013). In a dominantly qualitative implementation (Creswell, 2016), we used quantitative analysis to explore groups and relationships in the data that were subsequently complemented and contextualized through qualitative analysis. We are thus able to integrate diverse data sources and derive insights from the interplay of scale and richness of the available empirical material.

The analytical approach was highly iterative and consisted of multiple rounds of analysis and re-sampling. Similar to what others have termed ‘exploratory data loops’ (Eck and Uebernickel, 2016), the analysis moved back and forth between available empirical material, emergent constructs and relationships, as well as extant literature. In retrospective, two stages were crucial during the exploration and as such frame the analysis; (i) purposeful sampling, (ii) sub-setting. Proceeding through the analysis, each stage aimed to create “bridge assumptions” (Kelle, 2015) in order to pass on more and more refined ideas to subsequent stages. Both stages prepare the exploration of the empirical material by reducing complexity and dimensionality presented in the large scale data sets. The following section describes these stages.

Purposeful sampling

A first bridge assumption guiding our analysis was that of a situation in which the boundary resources of the platform become unstable. This presented a “sampling frame” (Kelle 2015; p. 602) for purposeful sampling (Yin, 2015). This approach often guides mixed-methods designs as it enables to sample according to emergent conceptions about a phenomenon of interest (Creswell, 2016; Kelle, 2015). As such, we assumed to find elevated levels of coordination work on the platform during times when boundary resource interfaces are changing through re-negotiation (Eaton *et al.*, 2015), or operator’s policy change (Karhu *et al.*, 2018). In short, in situations when stable, standardised interfaces deviate from standardisation and become unstable.

Therefore, we used our experience of working with the OpenStreetMap platform over the last four years to identify two episodes of platform boundary resource instability.

The first episode of boundary resource instability is the introduction of version six (“06”) of the OSM web API used to retrieve, edit, and store geo-data. The change consisted of the deprecation of an earlier API version in favour of a redesigned data model and data operations available via the RESTful web API (Ramm *et al.*, 2010). The API is crucial for the seamless interaction of participants with the platform. As one participant points out:

“Because there is an open API, anyone can come along and write an editor. Anyone can come along and write a renderer. Because it is not a great big monolith, then improvements can happen in isolation without getting anyone’s permission. And sometimes they just come from anywhere.” – Informant interview (secondary) in Coast (2015; p.123)

The new API version imposed changes to the established modes of operation of every participant type. For instance, every participant using the API to query OSM data for their own web services needed to adjust dependencies and align with the new functionality. Discussion in the community were abundant and we use the time from 2007 to the actual change over date in April 2009⁴⁵ to track communication and development contributions during that time.

The second episode of boundary resource instability was the change-over to an Open Database License (ODbL). The platform was initially launched with a CC-BY-SA 2.0 license – a common framework for sharing creative work, but unsuitable for the use, and re-use of data⁴⁶. As a consequence, it was decided to adopt a license in accordance with the Open Data Commons (ODC) framework. This allows free data use and re-use as long as the using party grants the same rights to others⁴⁷. The change over to the new license affected the logic of how and which geo-data is available for future re-use⁴⁸. The change commenced in July 2010 after which every newly-joined user was automatically signed up to the new license agreement. All existing participants (users, developers, external services, etc.) had to opt in to the new agreement. From September 2012 onwards the new license was intact and geo-data points that were not certified for re-use were automatically redacted at this point. To prevent circulation of the respective data, non-compliance with the new agreement would result in deletion of data by that user and potentially any derivative resource. This change necessitated

⁴⁵ https://wiki.openstreetmap.org/wiki/API_v0.6

⁴⁶ https://wiki.osmfoundation.org/wiki/Licence_and_Legal_FAQ/Why_CC_BY-SA_is_Unsuitable

⁴⁷ <http://opendatacommons.org/licenses/odbl/>

⁴⁸ https://wiki.osmfoundation.org/w/index.php?title=Licence/Historic/We_Are_Changing_The_Licence

substantial coordination efforts among platform participants. As one report summarises:

“Changing the license of a project as big as OpenStreetMap is, however, a very complex operation, because everyone who has ever contributed data needs to agree. If some contributors don’t agree, or can’t be reached, then their data can’t be used after the license change. If too many refuse to agree, then the whole idea of change has to be scrapped” – quoted from Ch. 20 “License Issues When Using Data” in Ramm *et al.* (2010)

We focus on the time from 2009 to the actual change over date in May 2012 to study the communication and development taking place during the license change.

Both events present major shocks to the functioning of otherwise stable and standardised boundary resources deployed on the platform. As a result, platform participants are forced to react to ensure future interaction with the platform. Mappers need to understand their rights and obligations when working with geo-data. Developers need to implement new technical specifications and understand updated documentation. External application providers need to align their services with new modes of operation on the platform. All of these activities likely resulted in increased coordination effort as well as increased development work. More importantly, we were interested in whether times of boundary resource instability involve participants that are not regularly part of the ongoing coordination work on the platform. Our interest was in the question of whether participants are involved in the coordination work during these episodes that would not normally be members of a clear-cut core or periphery group of participants. As such, these episodes provide a stellar opportunity to deepen our understanding of alternative forms of participant coordination on platforms. The focus on a boundary resource that grants access to platforms resources (i.e. the API) and a boundary resource that shares the platform’s intellectual property rights (i.e. the license) meaningfully captures the two main ways platform scholars recognise to open platforms for participant contributions (see e.g., Karhu *et al.*, 2018).

Sub-Setting

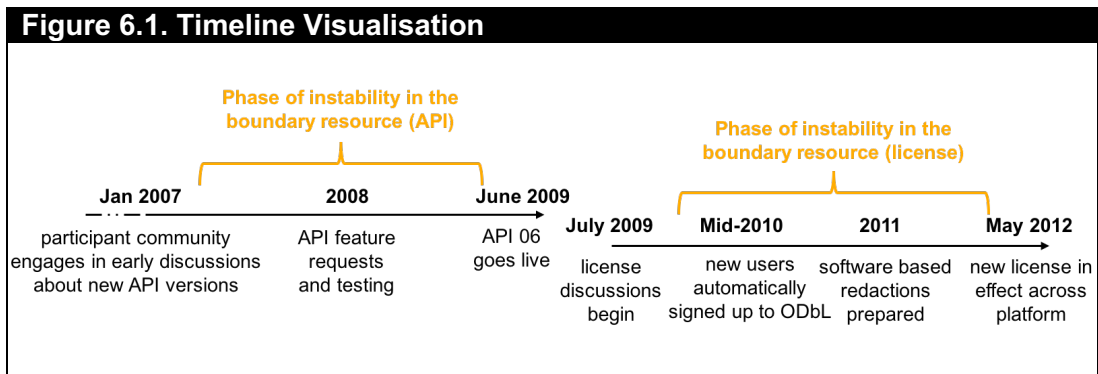
Yet another bridge assumption that guided our analysis is that of a ‘thematic subgroup’ in the communication among participants. To that end, each episode is constructed from the data by limiting the scope. We therefore filtered the communication and development data with the aim of creating subsets that reflect the communication and development contributions of participants pertinent to the respective boundary resource changes.

Form earlier work on the case we know that the relationship between participant contributions to communication and participant contributions to technical development are most promising on the levels of individuals. Specifically, the involvement of individual participants in topic subgroups of the e-mail communication.⁴⁹ As described above, e-mail communication unfolds in discussions threaded into topics. Most visibly, those topics are captured in subjects of e-mail messages. We therefore filtered the communication data and applied a keyword search on the e-mail subjects within the time windows framing the two respective episodes. The used keywords reflect concepts and terms relevant to the episode and were informed by interview partners and our experience from working with the case. We provide a full list of the keywords in the appendix.

Both episodes of boundary resource instability are well documented (Coast, 2015; Ramm *et al.*, 2010). While they span a substantial time window of the platform’s development, the episodes are temporally and thematically well separated⁵⁰. As such, the episodes present no overlap with each other. This reduces confounding bias in the analysis. Figure 6.1 below visualizes the episodes and key events over time.

⁴⁹In a pre-study, we found that using structural properties of the e-mail communication network for predictions of contributions to technical development leads to mixed results. Measures of technical development contributions (e.g., number of changes, or size of changes) and how well the contributions to participant communication predict that performance is contingent on the level of analysis. For instance, we found differences in the performance of subgroups vs. individuals. The pre-study was presented as “Hukal, P., (2018): Network Structure and Digital Platform Development: Showcasing the Generation of Research Questions through Inductive Analysis of Trace Data, *Academy of Management Specialized Conference: Big Data and Managing in a Digital Economy*, 18-20 April 2018, Surrey, UK” and is enclosed as an addendum to this essay.

⁵⁰ In our analysis we only found one e-mail discussion that explicitly referred to both episodes: The subject line of that e-mail discussion was “API changes for license change”. However, the discussion only started in March 2012 and our temporal framing of this e-mail discussion thus made it easy to assign this occurrence to one and only one episode.



6.4. FINDINGS

Across and within the two sampled episodes, we explore participant coordination on digital platforms in times of boundary resource instability. The findings explicate (1) detailed dynamics of platform participation, and (2) point to a form of coordination hitherto overlooked in the literature. In this, we concentrate on coordination among participants, focusing on the link between communication amongst participants and their contributions to development work.

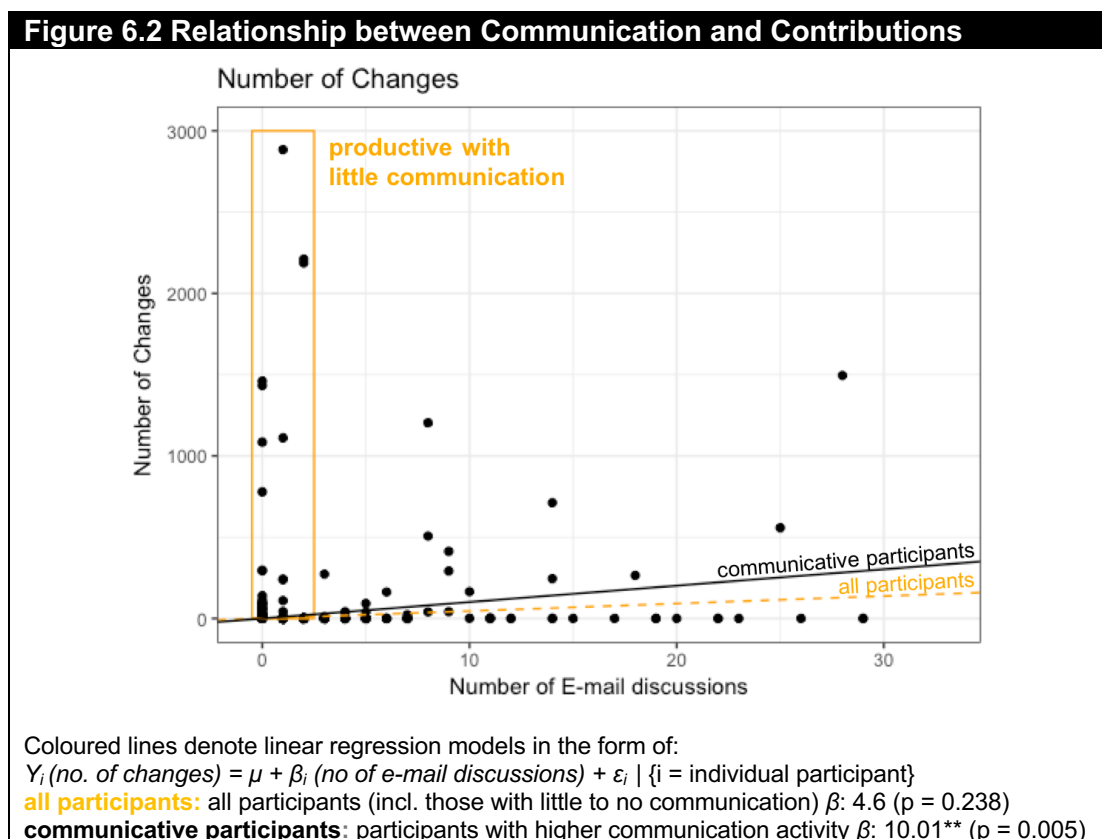
The findings are presented in two steps. First, we describe how a core-periphery expectation does not exhaustively capture activity among participants on the platform. We show that participants with very few interactions are nonetheless among the most prolific contributors. Second, exploring this counter-intuitive finding, we demonstrate how such participants go about coordinating their contributions by examining their e-mail communication in detail. As a result, we explain how rare, yet meaningful touchpoints with the core group of participants help coordinate others in times of boundary resource instability.

6.4.1. Deviation from Core-Periphery Coordination

Core members of a participant community are likely to engage in various aspects of a development – including coordination work (Dahlander and O’Mahoney, 2011). This implies that, if a clear core-periphery pattern existed, we would assume some degree of equivalency between participant involvement in communication and contributions to development. However, we find that, by and large, there is no such equivalency between the two aspects of participant engagement. This indicates that across the entire group of participants on the platform, a core-periphery pattern is not clearly

evident. Instead, we find a stratum of participants that communicates very little but is highly productive in terms of contributions to development work.

We visualise this finding in figure 6.2 below. The plot shows the number of e-mail discussions per participant in relation to the number of changes committed across both episodes of boundary resource instability. Visual inspection reveals a group that does not fit a clear core-periphery expectation (orange rectangle). This group commits a lot of source code changes, but does not engage in the communication that informs parts of the technical development. The expected core-periphery pattern only holds when this group is excluded. Only then is the number of e-mail discussions a significant predictor for contributions by individual participants. We indicate this by the linear model predictions visualised by the solid and dashed lines. Only the subset without the observations in the orange rectangle, follows a linearly increasing pattern of statistical significance⁵¹.



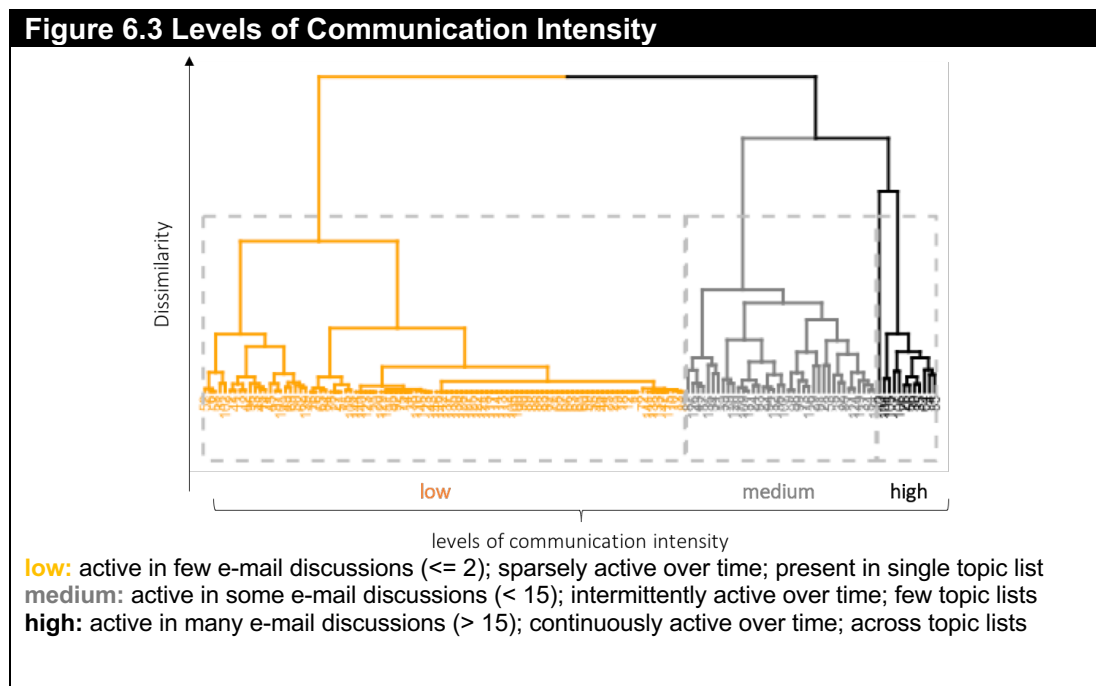
This first step of the analysis highlights that participation on the platform varies in a way that is unaccounted for when expecting a core-periphery pattern in coordination.

⁵¹ See the regression tables in appendix

We explored this divergence in more detail by delineating levels of engagement in the communication with others. We applied a set of clustering algorithms to determine potential partitions in the intensity with which participants engage in communication. To that end, we used the following metrics to cluster participants according to their activity in e-mail communication during both episodes:

- (i) the number of e-mail discussions a participant is engaged in,
- (ii) the average time between e-mail messages,
- (iii) the earliest and latest time points of e-mail discussions during the episodes.

Indeed, in both episodes the cluster analysis indicates a separation of the two datasets in three groups⁵². The optimal number of clusters was determined by majority rule of independent runs of clustering techniques⁵³ using 25 indices for optimisation (Charrad *et al.*, 2014). The three groups are characterised by the intensity with which participants engaged in e-mail discussions during the two respective episodes.



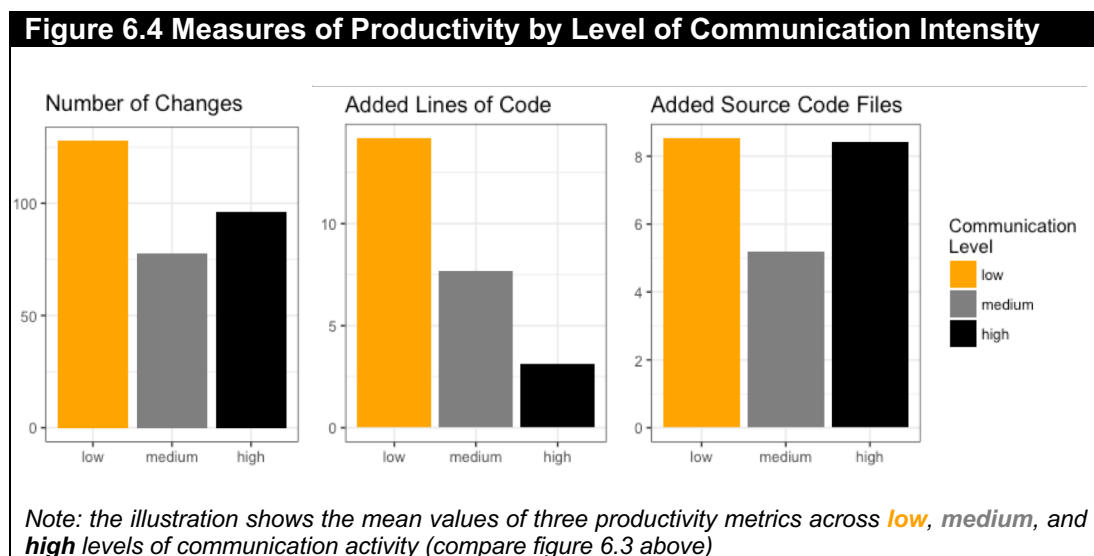
In the above cluster dendrogram, line endpoints denote participants on the platform that were active in both communication and development during the license change episode. Vertical distance represents dissimilarity between participants. Similarity is

⁵² Groups were formed with the constraint of having to include more than one observation.

⁵³ All clustering techniques are based on a matrix of squared dissimilarity scores in Euclidian space. For details and all indices used in the analysis see: Charrad M., Ghazzali N., Boiteau V., Niknafs A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set., *Journal of Statistical Software*, 61(6), 1-36

based on participant communication engagement (i.e. number e-mail discussions, number of topic lists, and activity over time). A higher distance until vertical lines converge is thus proportional to a higher dissimilarity between two participants' communication intensity. Colour codes visualize groups of participants that are more similar to each other than they are to the other groups. We include verbal descriptions of the found partitions below the illustration in figure 6.3.

These groups of varying levels of communication intensity are meaningful as they underscore the deviation of participation from core-periphery patterns. For instance, we contrast the level of communication intensity with the work each participant contributed to the technical development in the same time period. In figure 6.4 below we visualize this divergence by plotting the mean values of different measures of productivity across levels of engagement in communication (here during the time of the license change; 2009-2012).



This confirms the initial finding that a group of participants that rarely engages in communication still scores high across a variety of productivity measures⁵⁴ (here; number of changes, number lines of code changed, number of source files added).

6.4.2. Coordination among participants

⁵⁴ We find similar, albeit less pronounced, patterns in the API change episode. Here, participants with low communication intensity still contribute more to development than the group with moderate levels of communications (Wilcoxon-Rank Sum Test of the mean of committed changes: $W = 733.5$, $p\text{-value} = 0.04777$). And even if the group of low communication intensity is not the most productive, it is remarkable that no difference exists with participants of high levels of communication. E.g., in terms of lines of code changed: ($W = 360$, $p\text{-value} = 0.7875$) indicating no difference between the two groups.

The insight that a productive yet rarely communicative group of participants exists, motivated the investigation of the content of the e-mail discussions that such participants engage in. We therefore retrieved the e-mail discussions that involved these participants and examined the respective messages in full. In total, we analysed 20 e-mail discussions consisting of 235 e-mail messages.

In an open-coding content analysis (Schreier, 2014), we first read all e-mail messages and took note of generic attributes of the discussions with regards to coordination activity (Malone, 1988; Malone and Crowston, 1991). We mainly focused on what has been said, by whom, which information is being exchanged, and what – if any – outcome was achieved. In a second round of reading, we derived higher order constructs by aggregating the generic attributes into shared characteristics across the discussions. Based on these characteristics, we were able to discern patterns across the e-mail discussions.

Particularly, we found that the e-mail discussions shared salient attributes across them. We find that if participants with no clear fit in a core-periphery distinction engage, they are involved predominantly in highly detailed, collaborative, and constructive discussions. We found that 19 out of 20 e-mail discussions in our sample shared the characteristics we present below. Three e-mail discussions were off topic yet two nonetheless share the characteristics. Table 6.2 below provides a full list of the e-mail discussions we examined in detail.

The first characteristic we noted was the *'level of detail'* provided in these discussions. Shared details concerned technical specifications of the boundary resource and its change process in general. In most e-mail discussions we discerned two ways in which details were exchanged; (i) the explanation of a subject matter in great detail, and (ii) the elaboration on details in a subsequent message, either in discussions or by an additional participant providing further information. In the e-mail discussions that were on topic, 15 out of the 17 discussion threads presented a pattern of explaining and elaborating on details in this way.

For instance, one participant explains undocumented functionality in the new API used by other platform participants:

“One of the things API 0.6 will require is that DELETE requests can have a payload. While the documentation states it nowhere, at least the Sun Java implementation doesn't allow it. There's a hack in place to make it work, but

*now it will *only* work on the Sun Java implementation, which sucks. Now, there is the apache Http Client which does everything we want, but it requires some 400k of JAR files.”* – participant e-mail discussion “getting josm to do 06 properly” March 2008

In another discussion about data re-use, a participant elaborates on an earlier mention of the intricacies of the license with one editing tool used by other participants:

“Compiling is not illegal when non-illegal parts are integrated. What may be illegal is releasing a compiled version when declared with wrong license. Nevertheless, for distributions the result is the same :-). JOSM's [a software tool] code is stated as GPL V2 or later. If Debian requires explicit specification, then you can savely [sic!] declare it as GPLv3. There is no necessity [sic!] for us to do so with the codebase, as the ANT parts aren't directly included in the source repositories.” – participant e-mail discussion “possible copyright violation” July 2009

Another characteristic that these e-mail discussions shared is what we summarise as ‘*spanning sub-groups*’. It was noticeable that the e-mail discussions tied together members of different groups of platform participants. This meant either including participants from other topic lists, pointing to participants known to engage in specific technical developments, or simply copying in information from other lists or forums into the discussions.

In preparation of the API change and the many technical developments that needed to be coordinated, one participant directly refers to other participants by copying them into the relevant discussions:

“[...] I'll set up a segment free OSM database on dev, consisting of linear ways made up of ordered list of nodes, with direction (no superways as yet!) and based on planet.osm. It could also have its own API ([CC: user] - maybe this could tie in with Potlatch? [a software tool])” – participant e-mail discussion “segment free api and database” Jan 2007 [user name omitted]

Similarly, one participant ties in members of other domains in this message during the license change episode:

“I've started emailing the authors some time ago. The main contributors ([user1] and [user2]) would license their work under the terms of CC0. So do the following committers: [user3, user4, user5, user6]. A few people I haven't contacted so far” – participant e-mail discussion “License of Mapnik Style” April 2012 [user names omitted]

Yet another characteristic that was pertinent in these discussions was the formulation of future course of action. We refer to these characteristics collectively as ‘*charting a way forward*’. We found that the suggestion of actions came in two salient forms; either as announcement to others, or as collective fact finding. For instance, a participant announced that changes have been implemented and suggested how others can align with the new functionality.

During the license change, one participant announces what is expected in future development and how it impacts other software tools used on the platform:

“I’ve made a Wiki page that details the expected changes in API behaviour as a consequence of the license change:

http://wiki.openstreetmap.org/wiki/Open_Database_License/Changes_in_the_API

There’s no testing API available yet because the changes haven’t been coded yet.

JOSM will probably be affected in two places - one where it downloads an object history, and the other is where it downloads an old changeset and/or old versions of objects in order to try and revert the changeset.”
– participant e-mail discussion “API changes for license change”, March 2012

We often found e-mail announcements such as this, regularly appearing towards the end of discussions. Typically, these announcements include measures being taking by participants and communicating them clearly so others can prepare for the changes going forward.

Alternatively, participants converged on a solution through discussion and agreed on actions to be taken. For example, upon finding an error in a first API 06 implementation test, one participant summarises the solutions found:

“The problem was that Merkaartor [a software tool] (actually Qt) first sent unidentificated [sic!] requests, and only sent identificated [sic!] ones upon receiving a 401 from the server. [...] I will modify Merkaartor to always send indentificated [sic!] requests, which both solve the problem and will reduce the requests to the server. Win-Win, then :-) Thanks for the assistance”
– participant e-mail discussion “crowd sourced testing of api 06” December 2008

Lastly, we found that most of the e-mail discussions included members of the participant group that would traditionally be regarded as core. That is, these discussions included participants that are very active in both, communication and development, work. These participants contributed to e-mail threads by providing details, expanding or clarifying, or formulating a course of action. All but one of the 17 focal e-mail discussions had a member from the core participant group represented.

Table 6.2 below summarises the characteristics of the e-mail discussions in this step of the analysis by highlighting the aggregate categories used to characterise e-mail discussions among participants.

Table 6.2 Overview of E-Mail Discussions and Characteristics							
		Explain details	Elaborate details	Span subgroups	Chart way forward	Core member	
Episode	Subject of e-mail discussion	Characteristics					Time
API Change	segment-free database and api	x	x	x	x	x	Jan 2007
	Essen datamodel paper	offtopic					Apr 2007
	getting josm to do 06 properly	x	x		x	x	May 2008
	0.6 api hack-a-thon date	x	x	x	x	x	Oct 2008
	crowd sourced testing of osm 0.6 api	x	x	x	x	x	Dec 2008
	relation editor in 06	x			x	x	Jan 2009
	rantings about api 0.6	x	x	x	x	x	Feb 2009
	api 06 delete question	x	x	x	x	x	May 2009
	json/geojson output format for 0.6 api	x	x		x	x	Jun 2009
License Change	deletion conflicts/conflict resolution*	x		x	x	x	Jun 2009
	possible copyright violation	x	x	x	x	x	Jul 2009
	api/already deleted nodes one at a time**	offtopic					Aug 2009
	licensing question	x	x	x	x	x	Aug 2010
	nearmap community license	x	x	x	x		Aug 2010
	Status of pre and post quake	offtopic					Aug 2010
	odbl	x	x	x	x	x	Jan 2011
	license change plugin	x	x	x	x	x	Jul 2011
	API changes for license change	x	x	x	x	x	Mar 2012
	License of Mapnik Style	x	x	x	x	x	Apr 2012

In summary, we found that participants with little engagement in communication, yet numerous contributions, are involved in the e-mail discussions that are detailed, collaborative, and constructive for future work on the platform.

The results of this analysis indicate three findings about the dynamics of coordination on digital platforms in times of boundary resource instability:

- First, across all participants on the platform, no equivalency between communication and development work exists, indicating that a clear core-periphery pattern is not exhaustively representative of the coordination going on during times of boundary resource instability.
- Second, participants that do not fit the core-periphery distinction are nonetheless highly productive in terms of their contributions to technical development.

- Third, when such participants engage, they engage in meaningful, rare, yet constructive touchpoints with other productive members that would be considered members of a core group.

6.5. THREATS TO VALIDITY

Before discussing our findings, we will reflect on aspects that may threaten the validity of the conclusions drawn. First of all, one might be concerned with measurement errors biasing the analysis. Chiefly among them is the concern about the available data from the e-mail archive and its ability to capture coordination among platform participants. For instance, the common internet relay chat (IRC) protocols are used among OpenStreetMap platform participants to discuss issues. However, this source is not available for data collection. IRC works in real time and is not automatically recorded or saved to an archive. This makes such chats irretrievable in our empirical setting. Although we acknowledge that other forms of communication are popular among developers, we have reason to be confident in the adequacy of the used data from the e-mail list. For instance, the OSM community forum states:

“Email-based mailing lists are probably the main communication channel for the community” – entry on OpenStreetMap Wiki Forum [emphasis in original]

Furthermore, an archived discussion among early contributors to the platform argues for the preference given to forum and e-mail list as means of communication:

“[...] IRC is fine when many users are online, but you cannot easily post a question and check back a day later to see if anyone bothered to answer. [...]. There are a lot of people preferring forums over mailing lists and vice versa, so why not give them a choice. After all it is possible to setup a mailing list <-> forum gateway so you won't have to miss anything”
– participant discussion on OpenStreetMap Wiki archive

This assures us of the adequacy of a main source for our data collection.

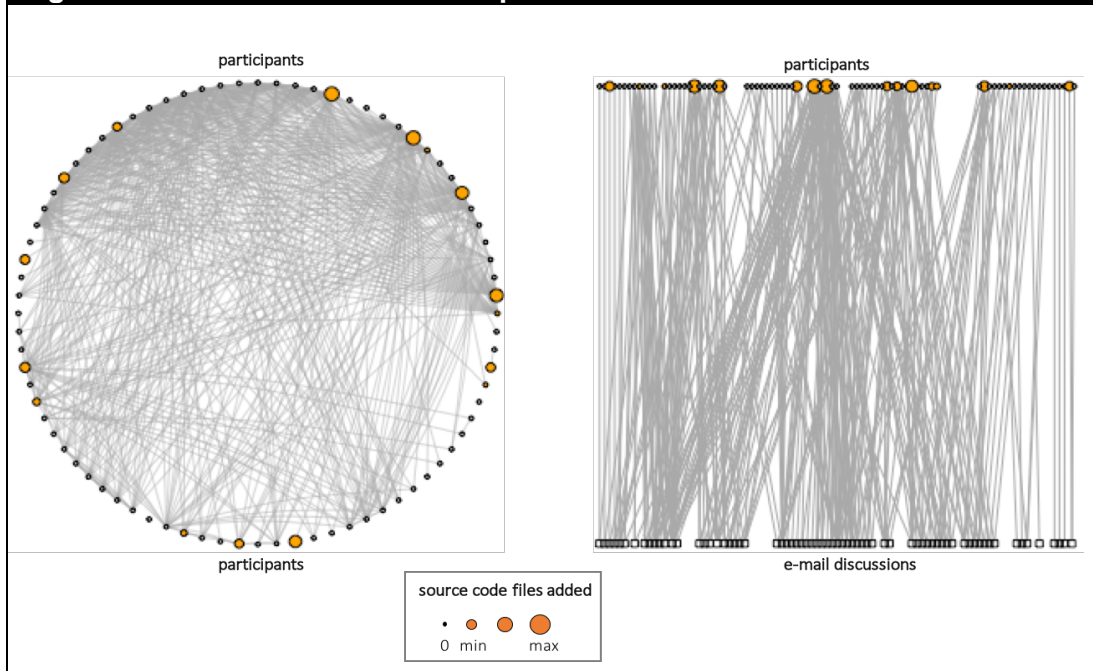
Further reduction of measurement error hinges on the correct identification of platform participants in the data. This was necessary to relate activity in communication with activity in development work. By cross-referencing individual participants in both data sources, we can confirm a high level of accuracy. The platform participants we could identify in the communication data account for the vast majority of all source code changes committed in each episode that we studied; ~92% during the license

change; ~84% during the API change over. This number further increases to ~98% in both episodes if developers are excluded that never participate in the e-mail lists. This assures us of the efficacy of our analysis as we can account for more than 9 out of 10 source code changes by tracing participants in communication and development logs.

Yet another source of bias is related to the functioning of GitHub as a main vehicle for development work on the platform. Often, projects on GitHub grant the rights necessary to commit changes to a selected group of contributors (Tsay *et al.*, 2014). This would bias our measures by inflating contributions of those participants with commit rights. To mitigate this, we use the information on authorship of source code changes instead of the actual committing. This minimizes measurement error further as we circumvent aspects such as who has commit rights and giving preference to the participants who created the initial change suggested for inclusion to the codebase.

Lastly, we checked for alternative explanations for the lack of fit between communication and development activity. It was important to us to rule out that activity in the development on the platform is contingent on structural or temporal aspects of participant's engagement in communication with others. Therefore, we applied network analysis and sequence analysis techniques to the communication and development activity data in both episodes. The result was that neither approach can reasonably predict the productivity of participants on the platform. That is, the structural position of a participant or the time point of engagement during an episode does not comprehensively explain how much a participant engages in development work. In figure 6.5 below we visualize participant involvement from a structural and a temporal perspective during the API 06 change episode.

Figure 6.5 Visualisations of Participant Coordination and Contribution



The left-hand side of the panel shows the interactions between participants in the e-mail discussions. Each point in the circular network projection represents an individual participant. Each line represents a case when two participants engaged in the same e-mail discussion. The size of each point is proportional to development work on the platform during the same time (here in terms of the number of added source code files). Clearly visible are differences in the communication and production levels. While some participants engage a lot in e-mail communication, others engage in fewer discussions. As highlighted in orange, the lack of engagement in communication in the mailing lists does not prescribe the amount of development contributions in the same time window. Some sparsely connected participants (few lines) are nonetheless highly productive (point size).

Similarly, the right-hand side of the panel shows platform participants' (top row) engagement in e-mail discussions (bottom row) over time (from left to right). Colour and size correspond to participants who contributed by adding source code files. Evidently, the level of such contributions does not seem to follow a temporal pattern, either. That is, when and how often, participants engaged does not prescribe their development contributions on the platform. Both in the beginning and in the end of the time window, as well as in connection to few and to many e-mail discussions, participants display varying levels of development contributions. We find similar

patterns in both episodes and across measures of productivity (lines of code, number of changes, number of added source code files).

In summary, we are confident about the surmised findings as important threats to their validity are accounted for in the design of the study.

6.6. DISCUSSION

In this study, we explored the dynamics of participant coordination on digital platforms during boundary resource instability. In so doing, we make a series of observations. First, we found that a stratum of participants exists that engages rarely, if ever, in coordination activity but is nonetheless highly productive in terms of contributions to the technical development. Second, when such participants engage in coordination work, they are involved in highly detailed, collaborative, and constructive information exchanges that include participants who would regularly be described as members of the core. These observations call into question the conventional, core-periphery view as expected in adjacent literature, as well as how we conceive of participant dynamics on digital platforms. Based on our findings, we therefore characterise a mode of coordination on digital platforms hitherto unaccounted for.

6.6.1. Pericentric Coordination

Our exploratory study uncovers a mode of coordination that has thus far been overlooked in the literature on digital platforms. In revealing a group that rarely, if ever, engages in coordination work with others, but still contributes substantially to the platform's operation, our study indicates that an additional form of coordination needs to be considered to capture coordination on digital platforms.

If the interface capabilities regulating interaction between the platform and its participants are changing, participants need to readjust their *modus operandi* when engaging with the platform. As participants exist that are very productive, yet hardly communicative, they do not fit core-periphery distinctions put forth in the literature on coordination (Crowston and Howison, 2006; Setia *et al.*, 2012). They are neither core, nor periphery. We would expect prolific participants to engage in coordination work at least to some extent. Indeed, past work has extensively dealt with the patterns

of contributors' progression "to the core" (e.g., Dahlander and Frederiksson, 2012) and the activity that such participants normally take on. For instance, core participants could be expected to engage in forums, chats, conferences, or as we expected, in the e-mail discussions with other participants.

The absence of such behaviour on digital platforms is thereby perfectly plausible. Under normal circumstances, coordination effort would be limited. After all, an important tenet of digital platforms is that contributions to the platform can unfold without coordination, simply by relying on standardised and stable interfaces (Yoo *et al.*, 2010) – such as boundary resources (Ghazawneh and Henfridsson, 2013; Eaton *et al.*, 2015). As a consequence, participants that contribute a lot to the platform's operation would not necessarily need to engage in coordination.

However, every now and then situations arise that make coordination necessary and as such induce collaboration and information exchange among participants. For instance, this occurs when boundary resources are subject to change and suffer from instability. Under these circumstances, platform participants can no longer rely on the stable interfaces provided by digital platforms that enable participants to contribute without coordinating their work. As a result, participants need to engage in coordination work in order to chart out a way forward that ensures a continuation of their individual contribution efforts.

How such coordination occurs is, however, unusual and has thus far not been addressed. The very low levels of communication that some participants exhibit could mean indifference. However, through our in-depth analysis of the communication by those participants we rule out a lack of interest as an explanation. To the contrary, the kind of communication these participants engage in implies a vested interest in the smooth operation of the platform. An interest by these participants is also indicated by the high levels of their contributions to the platform, in our case its technical development. In addition, these participants seek the exchange with members of the core, engaging in substantial and constructive coordination.

We refer to this dynamic as "pericentric coordination". Akin to a satellite in orbit that reaches a point closest to earth (known as a pericenter), pericentric coordination among participants on platforms includes participants that engage very little in

coordination. These participants enter coordination activity with other members before disengaging again and returning to productive work. The notion of a “pericenter” therefore adequately describes the recurrent situations in which participants engage in coordination with each other before continuing on their respective paths away from regular coordination activities.

The analogy of a satellite in orbit has implications for our understanding of coordination on platforms and how one can understand a mode of coordination that is pericentric. Naively approaching digital platforms with the expectation to find participant dynamics akin to what is established in adjacent literature risks misrepresenting what is going on on digital platforms. We observed participant engagement that does not neatly fit a core-periphery distinction as expected when extending the open source literature to digital platforms. These participants are highly productive – to the point where one could easily refer to them as ‘core’ contributors (Crowston and Howison, 2006; Setia *et al.*, 2012). At the same time however, these participants engage little – if at all – in coordination work with others. This leads us to our first proposition about this mode of coordination:

Proposition 1: Pericentric coordination brings together participants from inside and outside core-periphery groups of participants.

Coordination in the sense of this essay entails information exchange between two or more participants that neither participant alone would have to engage in individual work (Malone and Crowston, 1991). As is evident from our analysis, such participant coordination occurs through engagement in rare touchpoints distributed across time. Nonetheless, these coordination activities share salient characteristics that can be summarised as collaborative, constructive, and informative. As such, they are meaningful enough to satisfy the information requirements of individual participants. Consequently, as these touchpoints are rare yet meaningful, they do not occur on an everyday basis. This is underlined by the small number of e-mail discussions on topics that should result in heightened efforts in coordination and development. We surmise that these touchpoints are meaningful due to the level of detail they provide and the information exchange they present between participants. By sharing detailed and constructive information participants are enabled to return to their individual contributions without the need to reiterate shortly after. This substitutes for the lack of stability in the boundary resource that would normally facilitate interaction among

them. It does so in a way that ensures that participants can continue to contribute to the platform's operation once pericentric coordination has occurred. This leads us to our second proposition:

Proposition 2: Pericentric coordination occurs sporadically across time.

The main characteristics of pericentric coordination are not their temporal occurrence. What makes this form of coordination meaningful for platform participants is that these are the touchpoints that drive contribution to the platform. As indicated by our findings, neither the level of communication, nor structural position of participants, nor their temporal engagement determines contributions to development work. Across all participants, none of these measures alone predicts how much a participant will contribute to the platform. Instead, we surmise that the involvement of participants in crucial touchpoints with the core are determinative of contributions. Contributors cannot know how all the pieces fit together. Therefore, participants need to engage in information exchange at some point. What we describe as pericentric coordination brings participants to the fore that would not regularly engage. Yet, in detailed information exchanges with core members, these participants gain the knowledge necessary to proceed with their work and continue their contributions. These touchpoints of coordination work are crucial moments of interactions at which platform participants examine, reflect, and articulate the needs for further work in times when the boundary resource cannot reliably provide similar clarity. Such interactions help participants to take action individually and return to development work. In the content analysis of the e-mail discussions we have found that groups of participants often converge in understanding and chart out a way forward that is relevant to their respective area of contribution. This indicates that information is shared which allows the network of contributing participants to gain and share knowledge necessary to get on with their work. This line of reasoning informs our third proposition about pericentric coordination:

Proposition 3: Pericentric coordination positively affects contributions from involved participants.

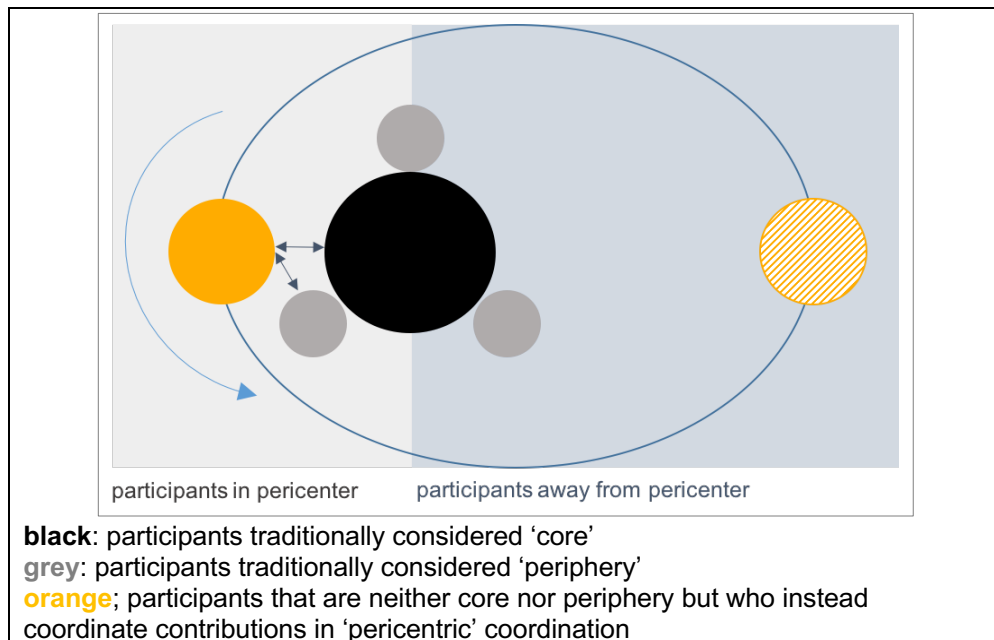
It is clear from our data that core-periphery coordination is taking place on the platform. However, it is not the only form of coordination. Our analysis is anchored in the situation of boundary resource instability. We decided to frame the study around

such extreme situations to explore fringe forms of coordination on platforms. The coordination activities we have observed would normally be subsumed by functionality of stable and standardised boundary resources. As a result, we conjecture that this form of coordination appears rarely. This is, however, not to say that we expect pericentric coordination only during times when boundary resources are unstable. Neither do we imply that these events are necessarily detrimental to the platform as a whole. After all, the changes in the boundary resources we focused on here were successful in the long run despite some instability during their implementation. The characteristics of pericentric coordination that we outline above allow to speculate about a number of situations occurring on platforms that would bring such coordination to the fore. For instance, it is plausible that the coordination occurs in a pericenter when new participants start to engage, when the platform enters a new market and hence invites new complementary contributions, or when contributors seek to understand changed technical functionality. All of these situations represent deviations from a normal mode of operation. We therefore surmise that pericentric coordination is generally a function of change. Where core-periphery coordination provides stability and reliability by predictably dividing tasks by role and extent of individual contributions, pericentric coordination captures coordination activity out of the ordinary. This informs our fourth proposition:

Proposition 4: Pericentric coordination occurs in addition to other forms of coordination on platforms.

In figure 6.6 below we visualise our view on pericentric coordination on digital platforms. While core-periphery groups of participants exist, a group of participants engages in coordination akin to a pericenter. Every now and then these participants seek information exchanges and stand in close contact with more continuously engaged participants before returning to activity of their individual contributions.

Figure 6.6 Schematic Representation of Pericentric Coordination



6.7. CONCLUSION

In summary, this study makes important contributions to the extant literature on digital platforms. We have shown that even sporadically engaging participants appear in the group of heavy contributors. Structure and temporality of engagement in coordination with participants does hence not exhaustively explain participant contributions on platforms. Instead, whenever participants require additional information in order to proceed with their contributions, a mode of coordination unfolds that differs from core-periphery expectations put forth in adjacent literature. What we refer to as pericentric coordination is observable due to instability in the interfaces that would otherwise facilitate interaction between participants and the platform. In such situations, many otherwise disengaged participants are forced to engage in coordination before they can return to their contributions.

Our findings have implications for how we research and understand participant dynamics on digital platforms. Theoretically, our study contributes to our understanding of coordination on digital platforms. IS researchers have embraced the logic of embedding governance and coordination decisions in component design (e.g., Wareham *et al.*, 2014). Crucially, these views rest on the assumption of stable and standardised functionality of interface capabilities, such as boundary resources (Eaton *et al.*, 2015). The immanent instability of boundary resources marks a departure from the normal *modus operandi* of a platform which has an impact on how involved participants interact with the platform and with each other. By focussing on situations

in which such interfaces become unstable, we are able to unearth dynamics of coordination as they occur out of the ordinary. As a first step in that direction, we here reflected on participant dynamics where the information about future contributions are coordinated in light of boundary resource instability.

In so doing, our study alludes to the unusual forms of coordination and emergent collaboration that are typical on platforms (Lyytinen *et al.*, 2016). This is important because we do not understand how participant networks form, how they coordinate contributions, and what the consequences of such dynamics are. Scholars often assume that network formation presupposes technology development (Germonprez and Hovorka, 2013). However, our findings indicate that properties of the participant network itself are of secondary importance. Digital platforms thrive on contributions in fluid and spontaneously emerging patterns. This is an inherent driver of innovation on digital platforms. For scholars interested in digital innovations on platforms, it is a matter of course that such innovations unfold serendipitously, often aided by spontaneous collaboration (Lyytinen *et al.*, 2016). Here, the proposed view on pericentric coordination articulates one dynamic by which such network formation – and the information flow implied therein – occurs and how it impacts participant contributions on platforms. As such, our study offers insights that are important to our growing understanding of innovation when mediated by complex digital technology assemblages (Nambisan *et al.*, 2017).

Lastly, our study also has methodological implications. What is commonly known as the Simpson paradox⁵⁵, captures the notion that large scale analysis approaches need to be handled with care as results may vary across levels of abstraction. Based on our work, we call for caution in the application of computational and quantitative methods without proper reflection. Consider that heavy contributors on a platform would never show up if examined through the standard lens of structural analysis of networks of their communication. In simple terms, a participant that engages once or twice in the e-mail lists would not be considered interesting or relevant to the functioning of the platform. However, dismissing such participant profiles based on one metric fails to

⁵⁵ From Wikipedia: “Simpson's paradox, or the Yule–Simpson effect, is a phenomenon in probability and statistics, in which a trend appears in several different groups of data but disappears or reverses when these groups are combined.”; accessed 11-07-2018

capture what is special about them and can hence incur major bias to an analysis. If platforms are analysed as mere networks of participants, one risks missing crucial insights.

APPENDIX TO CHAPTER 6

A) Keywords used in content filter

The table below shows the keywords used to query the subject line of the time filtered email discussions for each episodes. Curly brackets denote any combination used with the string preceding the brackets.

Episode	Keywords
API Version 0.6	"api" + {"06", "new", "v6", "ver 6"; "version 6", "0.6", "changeset", "relation", "segment"; "data" + {"base", "model", "scheme";
License change	"licen-" {"-sing", "-ce", "-ses"} + {"data", "attribution", "contribution"; "odbl"; "open" + {"data", "license"; "public" + {"data", "license"; "copyright"; "redact"

B) Downloaded E-Mail lists

The following table shows the topic lists that have been downloaded from the archive e-mail discussions.

List Name	Description
Design	Discussion of map styles, visualisation, and web design
Dev	General OpenStreetMap developer discussion
Osmosis-dev	Osmosis development; osmosis is a tool to handle various forms of geo-data
JOSM-dev	JOSM developer mailing list: JOSM = Java OpenStreetMap the name of a java-based editor
Graphhopper	Graphhopper is a java library used for routing navigation and mapping applications
Rails-dev	The web development framework Ruby-on-Rails is used to implement the OSM frontend platform and website
Potlatch-dev	Potlatch is a browser based editor for OSM geo-data
Merkaartor	OpenStreetMap editor for Linux, macOS and Windows
Tile-serving	Discussion of tile serving stacks and its development
Geocoding	Discussion of geocoding
Historic	Explorations of an OpenStreetMap Approach to Historic Mapping
Hot	Humanitarian OpenStreetMap Team
Imports	Data imports to OSM
Mapcss	Discussion among MapCSS implementers and stylesheet authors
Routing	Discussion of routing computations and use
Taginfo-dev	Taginfo development discussions; tag-info is a participant project that keeps track of the metadata descriptions used on the platform
Talk-transit	Public transport/transit/shared taxi related topics
Tagging	Tag discussion, strategy and related tools
rebuild	Forum for technical discussion of process, tools and functionality for rebuilding the OSM database ready for ODbL change-over

C) Informant Interviews

The table gives an overview of the role and focus in the primary informant interviews as well as the mode with which they were conducted.

ID	Informant Role (Affiliation)	Interview Focus	Mode
#1	OSM Consultant / Developer (3 rd party application)	Technical Functionality (metadata, API)	Video chat
#2	GIS Researcher (major UK research university)	Geo-data operations	Video chat
#3	Developer (3 rd party application)	Technical Functionality (workflow)	Video chat/email
#4	Founder & CEO (3 rd party application)	Use of OSM API and data, interaction with community	Video chat/email
#5	OSM Consultant / Developer (freelance)	OSM projects	Video chat
#6	PostgreSQL Developer (3 rd party application)	Database functionality	Email
#7	OSM Developer, Author (freelance)	Development of add-on tools, work with community	Email
#8	OSM Developer, Researcher (major UK research university)	Geo-data dynamics, work with community	Email/in person
#9	Software Developer (3 rd party application)	Technical Functionality (frontend, web)	Video chat

D) Regression Tables

Variable	<i>Model 1: all participants</i>		<i>Model 2: communicative participants (>2)</i>	
	<i>Coef</i>	<i>SE</i>	<i>Coef</i>	<i>SE</i>
E-Mail Discussions	4.58	3.88	10.181**	3.56
Intercept	73.90**	25.46	-10.378	39.65
Multiple R2 (%)	0.5		8.9	
Observations	249		85	

*Notes. Significance codes: * p < 0.05, ** p < 0.01, *** p < 0.001*

ADDENDUM: PRE-STUDY

Network Structure and Digital Platform Development: Showcasing the Generation of Research Questions through Inductive Analysis of Trace Data

ABSTRACT

The link between the structure of developer networks and actual development of digital platforms is understudied. In this short paper, we showcase how an inductive analysis of digital trace data from the OpenStreetMap platform aids the generation of research questions with high relevance for the information systems field. We explore network structures and their effect on platform development on three levels; network, sub-network, and individual. Our findings indicate that the influence of structural properties of networks on platform development differs across levels of abstraction. This motivates future research to move beyond the analysis of networks in their own right and instead explore the relationship between network structure and platform development in more detail to understand interaction of social and technical aspects of digital platforms.

Introduction

Interaction with digital technology produces data of potential interest to scholars and practitioners. In conjuncture with impressive developments of computational tools for collection and analysis, these ‘digital traces’ enable information systems researchers to capture phenomena that were unobservable or even non-existing in the past (Gaskin *et al.*, 2014; Hedman *et al.*, 2013).

One of these phenomena is the relationship between the network of actors working on a technology and the technology’s actual development (Lyytinen *et al.*, 2016). However, network research in IS lags behind advances in the wider management literature (see Phelps, Heidl, & Wadhwa, 2012). As a result, the investigation of the relationship between network structure and an outcome of relevance to information systems research is nascent⁵⁶ and a link to platform development remains entirely underdeveloped. Past work concentrated on macro-level examinations of networks and its influence on development of technology as a whole (e.g., Amrit & Hillegersberg, 2010). However, development work on platforms is highly distributed and developers often self-organize in groups and single out parts of a technology to work on (Crowston *et al.*, 2007). Our understanding of networks of developers and their development work on a platform thus merits expansion. As ever more economic activity is rendered by

⁵⁶ Notable exceptions include: Baldwin, MacCormack, & Rusnak (2014); Gray, Parise, & Lyer (2011)

digital technology platforms, investigating the link between network structure and platform development is crucial (Lyytinen *et al.*, 2016).

In this paper we report an exploratory study to illustrate how an inductive analysis approach to trace data guides theorizing about digital platforms and generates further research questions. We define a digital platform as the extensible code basis of a technical artefact whose architecture enables complementarities to interact with the platform core (Gawer, 2014; de Reuver *et al.*, 2018). We conducted a multi-level analysis of a developer network and its relation to platform developments. The study has implications for anyone interested in big data analysis and digital traces in the context of digital platforms. Guided by minimal assumptions about the process underlying the data generation and by relying on measures of simplicity, it was our goal to generate classifications and reveal associations in trace data. As a result, we formulate research questions in a broad programme of information systems research; what is the relationship between network structures and platform development?

Data

We report data from the OpenStreetMap (OSM) platform – an open source geo-spatial data project. Referred to as the ‘Wikipedia of maps’ (Fox, 2012), the platform provides mapping data and related capabilities to third party developers. In this context, the paper draws from two data sources. First, the mailing list of the OSM developer community. The mailing list is a full archive of communication among 2575 developers and consists of >72,000 e-mail messages. From the email archive we derive a network whose structure consists of nodes (individual developers) and ties (the communication between developers). Second, the project’s repositories on GitHub holding ~59,000 source code changes to the 19 software components of the platform. Upon identification of developers in both data sources, the data allows to relate a latent network structure from the communication data to observations of how components of the platform were developed over time. We focus on the time between 06/2007 to 06/2017 as a period of steady platform activity. We use the statistical programming language *R* and the package *igraph* (Csárdi and Nepusz, 2006) for network analysis techniques.

Analysis

The analysis follows an inductive approach to generate research questions from trace data. Induction is characterised by iteration over available empirical material with the goal to produce meaningful insight from patterns and associations emerging from unlabelled data (e.g., Berente & Seidel, 2014). While not theory in and by itself, induction often generates new pathways for theorization. Induction is especially valuable in the context of novel phenomena captured by trace data as it allows the grounding in ‘actual settings and processes’ (Vaast and Walsham, 2013). To maximise insight, we explore the relationship between network structure and platform development on three levels of analysis; the network-level, sub-network level, and individual level. In the language of network analysis this is equivalent to analysing an entire graph, its sub-communities, and its nodes. On the network level, global metrics of the structure of the developer network are related to the general development of the platform. On the sub-network level, groups of developers are investigated towards the components of the platform they worked on. Finally, on the individual level, roles within the developer network are associated with individual development contributions.

On each level of analysis, we follow three steps. In step 1, we identify groups in measures of network structure. We do this by employing PAM clustering (Saxena et al. 2017) on selected measures of network structure. Next, in step 2 the cluster results are related to measures of platform development. The objective is to ascertain statistically significant differences between the found clusters with respect to platform development by means of non-parametric sample tests. Lastly, in step 3 we derive a relationship between measures of network structure and platform development. Specifically, a simple model is estimated with a measure of platform development as a binary outcome and measures of network structure as predictors. Predictors are derived from the preceding steps on each level of analysis. As is common for inductive approaches exploring trace data, iteration and constant calibration characterised our approach (Berente and Seidel, 2014; Eck and Uebernickel, 2016). For example, outcomes on one level of analysis informed the investigation on subsequent levels which in turn helped to revise earlier analysis steps.

Findings

In reporting the exploratory findings, we focus on the outcomes from the modelled relationships on each level of analysis. Details on measures for each step on each level of analysis can be found in Tables 1 & 2.

On the network level, as the overall network grows, network structures with low density are 10-20% more likely to be more productive in terms of the number of changes made to the platform. An interesting finding since it implies that on the aggregate level, spread – not density – of a network is what drives the amount of development work on the platform. This contrasts positive effects between connectedness of a network and development expected in the literature. Relations among nodes are generally assumed to be indicative of the intensity of information sharing which is thought of as benefiting knowledge work (e.g., Phelps et al., 2012).

On the sub-network level, productivity is no longer a significant difference among groups. Instead, as the overall network scatters, groups with strong ties internally yet weak ties to the rest of the network, are 2-20% more likely to work on more complex platform components. This implies that connectedness is meaningful on the sub-network level, albeit for different reason. Dense sub-groups⁵⁷ within the overall network focus on more complex components of the platform and equally so, on more comprehensive changes⁵⁸. This qualifies the above finding as it captures a different aspect of productivity; within connected subgroups more changes are not the most salient feature, but the complexity of the task and complexity of the artefact is determinative. Findings like these are attributed to a sub-network satisfying its information need internally with little or no need to branch out to the entire network to produce their output (e.g., Phelps et al., 2012).

On the individual level, and across their subgroups, developers with higher involvement in different parts of the network and with more prominent roles (i.e. having many, diverse, and crucial ties to others), are 16-29% more likely to contribute more to the code base of the platform. This complements the preceding findings, as for individual contributions, a connected structural role of the individual does matter. By

⁵⁷ Sub-Group identification is based on node betweenness (i.e. strong and weak ties between nodes and groups)

⁵⁸ Scattered groups score significantly higher on lines of code changed by every commit ($W=7808.5$, $p < 0.001$)

extension, this finding is consistent with expectations on distributed software development (e.g., Daniel & Stewart, 2016) and knowledge work generally (Phelps *et al.*, 2012). Given its distributed nature, online software collaboration is characterised by a small number of developers being very productive, while others contribute very little. Assuming equivalency of the network structure, it is of little surprise that the most productive developers are also the best connected in the network. What is interesting about this finding is that it confirms what the literature expected all along. More connections translate into higher development output of the network node. However, our analysis shows that this does not uniformly translate across all levels of abstraction but only holds on the most granular level; the individual.

The results across the levels of analysis illustrate the potential of inductive approaches to generate research questions from trace data. Considering the different insights highlights avenues for further research on the relationship between network structure and platform development. A fundamental assumption of network analysis is that a network's structure is a function of its evolution and as a consequence, macro level properties are mirrored on the micro level (Barabasi, 2009). However, our analysis indicates that the level of abstraction with which researchers examine structural properties of networks exposes variability of outcomes. This has implication for further research on networks and digital technology development. First and foremost, effects of network structure might not follow the assumption of scale-freedom. Hence associating structural properties of networks to technical development might reveal differences in kind contingent on the level of abstraction.

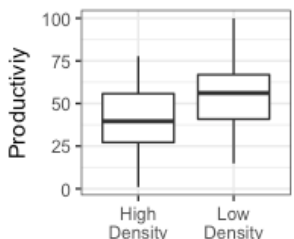
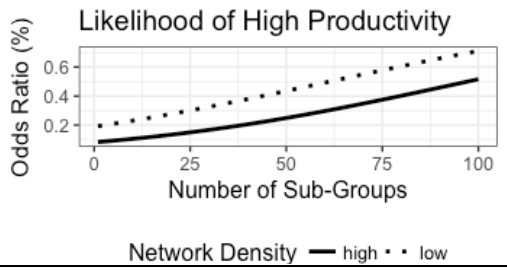
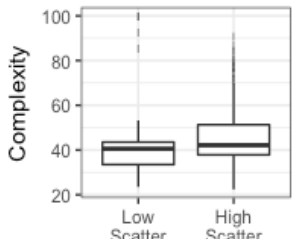
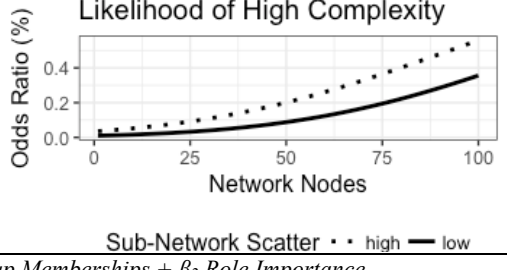
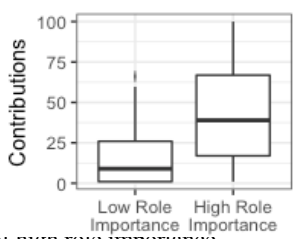
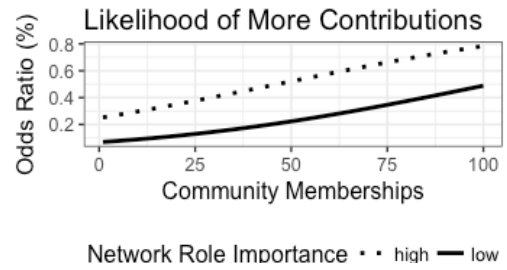
Further research is thus motivated by the desire to advance beyond the analysis of networks in their own right. Instead, exploring the relationship between networks and technology development in more detail promises to improve explanations of social and technical interaction. This guides the formulation of future research questions putting the diversity and dynamics of networks in centre of digital platform studies (cf. Lyytinen *et al.* 2016). One starting point could be the inconsistency indicated through the empirical analysis: How and why do structural properties of networks affect technology development? How and why do these effects differ across levels of analysis? How and why does network structure influence the trajectory of platform component development?

References

- Amrit, C., & Hillegersberg, J. van. 2010. Exploring the Impact of Socio-Technical Core- Periphery Structures in Open Source Software Development Exploring. *Information Technology*, 25(2): 216–229.
- Barabasi, A.-L. 2009. Scale-Free Networks: A Decade and Beyond. *Science*, 325(5939): 412–413.
- Berente, N., & Seidel, S. 2014. Big Data & Inductive Theory Development: Towards Computational Grounded Theory? *Proceedings of the Americas Conference on Information Systems*, 1–11.
- Crowston, K., Li, Q., Wei, K., Eseryel, U. Y., & Howison, J. 2007. Self-organization of teams for free/libre open source software development. *Information and Software Technology*, 49(6): 564–575.
- Csárdi, G., & Nepusz, T. 2006. The igraph software package for complex network research. *InterJournal Complex Systems*, 1695: 1–9.
- Daniel, S., & Stewart, K. 2016. Open source project success: Resource access, flow, and integration. *The Journal of Strategic Information Systems*, 25(3): 159–176.
- de Reuver, M., Sorensen, C., & Basole, R. C. 2017. The Digital Platform - A Research Agenda. *Journal of Information Technology*, 1–12.
- Eck, A., & Uebernickel, F. 2016. Reconstructing Open Source Software Ecosystems: Finding Structure in Digital Traces. *Proceedings of the 37th International Conference on Information Systems*. Dublin, Ireland.
- Fox, K. 2012, February 18. OpenStreetMap: “It’s the Wikipedia of maps.” *The Guardian*.
- Gaskin, J., Berente, N., Lyytinen, K., & Yoo, Y. 2014. Toward Generalizable Sociomaterial Inquiry: A Computational Approach for Zooming In and Out of Sociomaterial Routines. *MIS Quarterly*, 38(3): 849–871.
- Gawer, A. 2014. Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, 43(7): 1239–1249.
- Hedman, J., Srinivasan, N., & Lindgren, R. 2013. Digital Traces of Information Systems: Sociomateriality Made Researchable. *Proceedings of the 34th International Conference on Information Systems*, 1–17. Milan, Italy.
- Huang, S., Lv, T., Zhang, X., Yang, Y., Zheng, W., et al. 2014. Identifying Node Role in Social Network Based on Multiple Indicators. *PLoS ONE*, 9(8): e103733.
- Lyytinen, K., Yoo, Y., & Boland, R. J. 2016. Digital product innovation within four classes of innovation networks. *Information Systems Journal*, 47–75.
- Phelps, C., Heidl, R., & Wadhwa, A. 2012. Knowledge, Networks, and Knowledge Networks. *Journal of Management*, 38(4): 1115–1166.
- Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., et al. 2017. A review of clustering techniques and developments. *Neurocomputing*, 267: 664–681.
- Vaast, E., & Walsham, G. 2013. Grounded theorizing for electronically mediated social contexts. *European Journal of Information Systems*, 22(1): 9–25.

Table 1.

Measurement	Description
<i>Nodes</i>	The number of elements in a (sub-)network
<i>Group Membership</i>	Number of sub-group an individual engages with
<i>Distance</i>	The number of steps required to connect a given pair of nodes in a network
<i>Density</i>	Number of ties a node or (sub-)network has in relation to all possible ties
<i>Sub-Groups</i>	The number of groups within a network that have strong ties to each other but weak ties to other parts of the network
<i>Communities</i>	Similar to clusters yet ties are weighted by betweenness (the number of times a node lies on the optimal path to other nodes)
<i>Role Importance</i>	Compound measure of a node's number and criticality of ties as well as its adjacent nodes' centrality. see Huang et al. (2014)
<i>Productivity</i>	Sum of software changes per month
<i>Complexity</i>	Lines of code of a software component / or committed software change
<i>Contributions</i>	committing a change, starting a discussion, resolving an issue

	Data ¹	Step1: Clustering	Step2: Cluster Differences	Step 3: Relationship Derived from Clusters						
Network Level of Analysis	<p>120 monthly snapshots of the overall network structure and development of the platform as a whole</p>	<p>x: number of network sub-groups y: distance between nodes z: density of network</p> <p>Optimal clusters: 2 (size 48 / 72) Av. Silhouette Width: 0.46</p> <p>C1 (x/y/z): 49.3 / 48.9 / 68.4 C2 (x/y/z): 75.9 / 73.4 / 47.1</p> <p><u>Interpretation:</u> C1: high density C1: low density</p>	 <p>Cluster of less dense network is more productive (W = 2189, p = 0.001227)</p>	<p>$Y(\text{probit}) \sim \text{Productivity} > 50 = \beta_1 \text{ Number of Subgroups} + \beta_2 \text{ Network Density}$</p> <p><i>Productivity (SE)</i></p> <table border="1"> <tr> <td>Constant</td> <td>-1.397*** (0.473)</td> </tr> <tr> <td>Subgroups</td> <td>0.014** (0.59)</td> </tr> <tr> <td>Density: Low</td> <td>0.515** (0.247)</td> </tr> </table> <p>Obs: 120 LogLik: -76.053 AIC: 158.11 *p<0.1; **p<0.05; ***p<0.01 controls²; nodes, communities</p>  <p>Likelihood of High Productivity</p> <p>Network Density — high •• low</p>	Constant	-1.397*** (0.473)	Subgroups	0.014** (0.59)	Density: Low	0.515** (0.247)
Constant	-1.397*** (0.473)									
Subgroups	0.014** (0.59)									
Density: Low	0.515** (0.247)									
Sub-Network Level of Analysis	<p>217 sub-networks and their link to parts of the platform</p>	<p>x: number of nodes y: number of communities</p> <p>Optimal clusters: 2 (104 / 113) Av. Silhouette Width: 0.54</p> <p>C1 (x/y): 55.4 / 39.0 C2 (x/y): 75.9 / 71.3</p> <p><u>Interpretation:</u> C1: lowly scattered network C1: highly scattered network</p>	 <p>Cluster of scattered sub-network works on more complex products (W = 7083, p = 0.004512)</p>	<p>$Y(\text{probit}) \sim \text{Product Complexity} > 55 = \beta_1 \text{ Network Nodes} + \beta_2 \text{ Network Communities}$</p> <p><i>Complexity (SE)</i></p> <table border="1"> <tr> <td>Constant</td> <td>-2.330*** (0.578)</td> </tr> <tr> <td>Nodes</td> <td>0.019** (0.008)</td> </tr> <tr> <td>Scatter: High</td> <td>0.515* (0.306)</td> </tr> </table> <p>Obs: 217 LogLik: -93.58 AIC: 193.16 *p<0.1; **p<0.05; ***p<0.01 controls; density, clusters, distance</p>  <p>Likelihood of High Complexity</p> <p>Sub-Network Scatter •• high — low</p>	Constant	-2.330*** (0.578)	Nodes	0.019** (0.008)	Scatter: High	0.515* (0.306)
Constant	-2.330*** (0.578)									
Nodes	0.019** (0.008)									
Scatter: High	0.515* (0.306)									
Individual Level of Analysis	<p>135 identified developers and their contribution to platform development</p>	<p>x: number of group memberships y: network role importance</p> <p>Optimal clusters³: 2 (size 62 / 73) Av. Silhouette Width: 0.51</p> <p>C1 (x/y): 22.1 / 1.0 C2 (x/y): 64.4 / 50.5</p> <p><u>Interpretation:</u> C1: low importance C1: high importance</p>	 <p>Cluster of high role importance contributes more to product dev. (W = 3325, p = 0.00001145)</p>	<p>$Y(\text{probit}) \sim \text{Contributions} > 60 = \beta_1 \text{ Group Memberships} + \beta_2 \text{ Role Importance}$</p> <p><i>Contributions (SE)</i></p> <table border="1"> <tr> <td>Constant</td> <td>-1.121*** (0.239)</td> </tr> <tr> <td>Memberships</td> <td>0.015** (0.006)</td> </tr> <tr> <td>Importance H</td> <td>0.542** (0.215)</td> </tr> </table> <p>Obs: 135 LogLik: -61.275 AIC: 129.27 *p<0.1; **p<0.05; ***p<0.01 controls; rate of change, lines of code changed</p>  <p>Likelihood of More Contributions</p> <p>Network Role Importance •• high — low</p>	Constant	-1.121*** (0.239)	Memberships	0.015** (0.006)	Importance H	0.542** (0.215)
Constant	-1.121*** (0.239)									
Memberships	0.015** (0.006)									
Importance H	0.542** (0.215)									

¹ All data is log-transformed and rescaled to values of 1 to 100 for ease of comparison

² Results of control models omitted due to space limitations; effects are consistent for listed control variables

³ Optimal number of clusters would have been 4, reduction to 2 clusters yielded similar results at the expense of silhouette width of 2.8%

CHAPTER 7 – SUMMARY

This dissertation presents three empirical studies on aspects of growth and innovation of digital platforms. In doing so, each study makes substantial and original contributions to the literature on digital platforms. In particular, the most salient findings from the empirical work of this dissertation align with the identified contribution targets outlined in the literature review. Chapter four offers a detailed empirical account of generativity. Chapter five reflects on the intellectual tradition of modularity that underpins most studies on digital technology design. Lastly, chapter six focuses on the coordination of participants as a function of the instability of digital platform components. This final section briefly summarises the findings in each study and articulates potential for future research.

In chapter four, the dissertation studied how and why endorsements influence generativity on digital platforms. Generativity is a salient dynamic of digital technology and determinative for the growth and innovation of digital platforms. The question of how generativity can be influenced is thereby crucial for platform operators, since participants on the platform cannot know where generativity is desired. The study addresses this information asymmetry and theorises about endorsements as measures of information sharing between the platform operator and participants. Specifically, the study demonstrates that endorsements follow strategic motives with the intention to signal where and which participant contributions are desired on the platform. In so doing, the study extends the existing platform literature with a detailed view on how generativity on digital platforms can be stimulated by the platform operator using endorsements. The study argues that endorsements are crucial for stimulating generativity given the correct identification and appropriate communication of desirable qualities. Done correctly, endorsements are signals that can increase the output of a platform in scale and scope and as such provide a powerful lever for platform operators to guide generative change.

By studying purposeful actions by the platform operator, the study opens up an exciting direction for future research on digital platforms. The suggestion that generativity can be stimulated, at least to an extent, by deliberately engaging with and

steering activity on the platform is promising. Future studies have a plethora of possibilities to explore what other courses of actions are fruitful for platform operators, regulators, or participants to bring about growth and innovation on platforms that are generative. Gathering, creating, and combining digitally stored information across actors results in novel configurations of digital services that are beyond initial design, central control, or deliberate planning by platform operators. Exploring actions that make such changes manageable is aligned with a growing interest in creating theories of generative information systems.

Chapter five makes a theoretical and empirical contribution to the literature on platform-based ecosystems by reflecting on modularity – one of the most important intellectual underpinnings of digital technology platforms. Modularity's core promise is to increase the ability of a technical system to change by hiding complexity. Yet, in order to foster change, system designs are expected to follow modular design principles, first and foremost demanding that dependencies between modules are kept to a minimum. The study contrasts this conventional view on modular systems with the understanding of digital technology and its unique attributes. The study argues that as a consequence of their underlying technological properties, digital platforms become unbounded systems and require little to no coordination to resolve interdependencies. The unbounded character of digital technologies does not constrain architectural designs to the same extent as classical modular systems. Participants in digital platform ecosystems often leverage the shared architectures of digital technologies to draw from capabilities developed elsewhere. Digital platforms thus benefit from swift innovations enabled through increased interaction across shared architectures. The study underlines the recognition that digital technology is different not only in degree but also in kind, suggesting that platform operators manage unboundedness to stimulate the pace of change of their platform.

This study is a first step for possible investigations into design and organisation of digital technology and the desire to reflect on long-standing theoretical underpinnings. A fruitful route for further research would, for instance, be the identification of the kind of dependencies that enable high paced change, innovation, and eventually generativity. Here, research in IS stands to gain from further work on the design and organisation of digital technology in complex ecologies of interactions. Research on the supra-organisational forms in which digital technology is arranged is still nascent.

A deeper understanding of the structure, architecture, and design of such digital technology ecosystems is likely to fuel research for years to come.

In the last empirical study, chapter six reports on a mode of coordination hitherto overlooked in the literature on digital platforms. Framed by two episodes of boundary resource instability, the study explores how participants coordinate their contributions in times when normally stable interfaces risk becoming unstable. One finding is that structure and temporality of coordination among participants does not exhaustively explain participant contributions on platforms. Instead, whenever participants require additional information in order to proceed with their contributions, a mode of coordination unfolds that has thus far not been accounted for. The study theorises this mode of coordination on digital platforms as “pericentric coordination”. Pericentric coordination is observable due to instability in the interfaces that would otherwise facilitate interaction between participants and the platform. In such situations, many otherwise disengaged participants are forced to engage in coordination before they can return to their contributions. As a complementary mode of coordination, the study characterises pericentric coordination as bringing together participants from within and beyond core-periphery groups, occurring sporadically over time, and benefiting contribution levels of the participants involved. Platforms normally can do without such coordination but require well defined and stable interfaces to do so. In times when boundary resources become unstable, platform participants face situations that necessitate coordination before returning to their individual contributions.

More work is needed to completely understand coordination on digital platforms. It is conceivable that platforms bring about coordination forms hitherto unknown based on the dynamics of heterogeneous and distributed parties interacting on platforms. For instance, coordination on platforms can differ based on the kind of contributions that participants make. Plausibly, participants contributing to a technical development engage in different coordination than participants on platforms that share content. Furthermore, the level of abstraction in studies of coordination on digital platforms is potentially meaningful. Major developments, such as a new API or a new license under which a platform operates, may have substantial implications for the vitality and vibrancy of the entire ecosystem of external application developers. Such changes essentially affect the lifeblood of digital platforms as every affiliated party needs to

adjust. The consequences of such major developments on digital platforms are neither well researched nor well understood⁵⁹, and thus motivate a direction for future work.

Collectively, the presented empirical studies speak to the contribution targets identified in the literature review. Through approaches that utilise computational research techniques in IS, the dissertation provides detailed empirical accounts of phenomena of relevance to anyone seeking to better understand growth and innovation of digital platforms.

The importance of the topics covered in this dissertation extend beyond the information systems field. This research on digital platforms is thus also motivated by what Amrit Tiwana calls “the information systems advantage”⁶⁰. Studies on topical information systems phenomena promise to inform other disciplines by opening up digital technology artifacts. As a discipline with a deep appreciation for, and competence in, matters of digital technology, information systems research has the chance to inform scholarly investigation beyond its own discipline. Studies on digital platforms can thus meaningfully deepen the understanding of technology-enabled dynamics prevalent in other fields. This has become more pronounced as the use and development of digital technology permeates all spheres of our lives (Yoo, 2010). Consequently, understanding digital platforms is relevant for phenomena such as that economic transactions on online marketplaces, the changing nature of work through ‘sharing economy’ platforms, or the study of behaviour rendered on social media platforms. Both within and beyond the field of information systems, scholars can benefit from an understanding of how digital platforms grow and generate digital innovations.

It is hoped this dissertation has contributed towards that goal.

⁵⁹ notable exceptions include Förderer *et al.* (2018), and Nagaraj (2017)

⁶⁰ At a recent talk at Temple University, PA (link to slides: <http://people.terry.uga.edu/tiwana/pdfs/misc/isr2017ppt.pdf>)

REFERENCES

- Aaltonen A, Seiler S. 2016. Cumulative Growth in User-Generated Content Production: Evidence from Wikipedia. *Management Science* **62**(7): 2054–2069.
- Adner R. 2017. Ecosystem as Structure: An Actionable Construct for Strategy. *Journal of Management* **43**(1): 39–58.
- 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**: 306–333.
- Adner R, Kapoor R. 2016. Innovation Ecosystems and the pace of substitution: Re-Examining Technology S-Curves. *Strategic Management Journal* **37**: 625–648.
- Agarwal R, Gupta AK, Kraut R. 2008. The Interplay Between Digital and Social Networks. *Information Systems Research* **19**(3): 243–252.
- Agarwal R, Tiwana A. 2015. Evolvable Systems: Through the Looking Glass of IS. *Information Systems Research* **26**(3): 473–479.
- Ahonen P. 2015. Institutionalizing Big Data methods in social and political research. *Big Data & Society* (July-December): 1–12.
- Alaimo C, Kallinikos J. 2017. Computing the Everyday: Social Media as Data Platforms. *The Information Society*. Taylor & Francis **33**(4): 1–43.
- Van Alstyne MW, Parker GG, Choudary SP. 2016. Pipelines, Platforms, and the New Rules of Strategy. *Harvard Business Review* **94**(APRIL): 14.
- Alvarez RM. 2016. *Computational Social Science - Discovery and Prediction*. Cambridge University Press: New York.
- Amirian P, Basiri A, Gales G, Winstanley A, McDONALD J. 2015. The Next Generation of Navigational Services Using OpenStreetMap Data: The Integration of Augmented Reality and Graph Database. In *OpenStreetMap in GIScience - Experiences, Research, and Applications*: 211–229.
- Amrit C, Hillegersberg J van. 2010. Exploring the Impact of Socio-Technical Core-Periphery Structures in Open Source Software Development. *Information Technology* **25**(2): 216–229.
- Anderson P. 1999. Perspective: Complexity Theory and Organization Science. *Organization Science* **10**(3): 216–232.
- Arsanjani JJ, Zipf A, Mooney P, Schauss A. 2015. Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets. In *OpenStreetMap in GIScience*, Arsanjani JJ, Zipf A, Mooney P, Heiblich M (eds). Springer Science+Business: Heidelberg/New York: 37–58.
- Balazsi L, Baltagi BH, Matyas L, Pus D. 2017. Random Effects. In *The Econometrics of Multi-Dimensional Panels: Theory and Applications*, Matyas L (ed). Springer: Cham: 35–67.
- Baldwin C, Clark KB. 1997. Managing in an Age of Modularity. *Harvard Business Review* (September-October): 84–93.
- Baldwin C, Clark KB. 2000. *Design Rules - the Power of Modularity*. Cambridge, MA: MIT Press. MIT Press: Cambridge, MA.
- Baldwin C, Clark KB. 2006. The Architecture of Participation: Does Code Architecture Mitigate Free Riding in the Open Source Development Model? *Management Science* **52**(7): 1116–1127.
- Baldwin C, MacCormack A, Rusnak J. 2014. Hidden structure: Using network methods to

- map system architecture. *Research Policy* **43**(8): 1381–1397.
- Baldwin C, Woodard CJ. 2009. The architecture of platforms: a unified view. In *Platforms, Markets and Innovation*, Gawer A (ed). Edward Elgar: Cheltenham: 19–44.
- Ballinger GA. 2004. Using Generalized Estimating Equations for Longitudinal Data Analysis. *Organizational Research Methods* **7**(2): 127–150.
- Bamberger PA. 2018. AMD—Clarifying What We Are about and Where We Are Going. *Academy of Management Discoveries* **4**(1): 1–10.
- Barabasi A-L. 2009. Scale-Free Networks: A Decade and Beyond. *Science* **325**(5939): 412–413.
- Baskerville R, Wood-Harper A. 1998. Diversity in information systems action research methods. *European Journal of Information Systems* **7**(2): 90–107.
- Benbya H, McKelvey B. 2006. Toward a complexity theory of information systems development. *Information Technology and People* **19**(1): 12–34.
- Benkler Y. 2006. *The Wealth of Networks - How Social Production Transforms Markets and Freedom*. Yale University Press: New Haven and London.
- Benlian A, Hilkert D, Hess T. 2015. How open is this platform ? The meaning and measurement of platform openness from the complementors ' perspective. *Journal of Information Technology* **30**(3): 209–228.
- Berente N, Seidel S. 2014. Big Data & Inductive Theory Development: Towards Computational Grounded Theory? *Proceedings of the Americas Conference on Information Systems* : 1–11.
- Berente N, Seidel S, Safadi H. 2018. Data-Driven Computationally-Intensive Theory Development. *Information Systems Research* : 1–34.
- Bergtold JS, Yeager EA, Featherstone AM. 2018. Inferences from logistic regression models in the presence of small samples, rare events, nonlinearity, and multicollinearity with observational data. *Journal of Applied Statistics* **45**(3): 528–546.
- Boland RJ, Lyytinen K, Yoo Y. 2007. Wakes of Innovation in Project Networks: The Case of Digital 3-D Representations in Architecture, Engineering, and Construction. *Organization Science* **18**(4): 631–647.
- Bolker B. 2008. *Ecological Models and Data in R*. Princeton University Press: Princeton.
- Bolker B. 2016. Dealing with quasi-models in R. *R Vignette*. Available at: <ftp://mirrors.xmu.edu.cn/CRAN/web/packages/bbmle/vignettes/quasi.pdf> [1 June 2016].
- Boudreau K. 2010. Open Platform Strategies and Innovation: Granting Access vs. Devolving Control. *Management Science* **56**(10): 1849–1872.
- Boudreau K. 2012. Let a Thousand Flowers Bloom? An Early Look at Large Numbers of Software App Developers and Patterns of Innovation. *Organization Science* **23**(5): 1409–1427.
- Boudreau K. 2017. Platform boundary choices & governance: Opening-up while still coordinating and orchestrating. In *Entrepreneurship, Innovation, and Platforms (Advances in Strategic Management)*, Furman J, Gawer A, Silverman BS, Stern S (eds). Emerald Group Publishing, 37: 227–297.
- Boudreau K, Jeppesen LB. 2015. Unpaid Crowd Complementors: The Platform Network Effect Mirage. *Strategic Management Journal* **36**: 1761–1777.
- Bowman C. 2015. The role of technology in the creation and capture of value. *Technology Analysis and Strategic Management* **27**(2): 237–248.
- Brusoni S, Prencipe A. 2013. The Organization of Innovation in Ecosystems: Problem

- Framing, Problem Solving, and Patterns of Coupling. In *Collaboration and Competition in Business Ecosystems*, Adner R, Oxley JE, Silverman BS (eds). Emerald Group Publishing: 167–194.
- Brynjolfsson E, Kemerer CF. 1996. Network Externalities in Microcomputer Software: An Econometric Analysis of the Spreadsheet Market. *Management Science* **42**(12): 1627–1647.
- Budhathoki NR, Haythornthwaite C. 2013. Motivation for Open Collaboration: Crowd and Community Models and the Case of OpenStreetMap. *American Behavioral Scientist* **57**(5): 548–575.
- Campagnolo D, Camuffo A. 2010. The concept of modularity in management studies: A literature review. *International Journal of Management Reviews* **12**: 259–283.
- Castro FG, Kellison JG, Boyd SJ, Kopak A. 2010. A Methodology for Conducting Integrative Mixed Methods Research and Data Analyses. *Journal of Mixed Methods Research* **4**(4): 342–360.
- Ceccagnoli M, Forman C, Huang P. 2012. Cocreation of Value in a Platform Ecosystem: The Case of Enterprise Software. *MIS Quarterly* **36**(1): 263–290.
- Cennamo C, Ozalp H, Kretschmer T. 2018. Platform Architecture and Quality Tradeoffs of Multihoming Complements. *Information Systems Research* : 1–18.
- Chapin N, Hale JE, Khan KM, Ramil JF, Tan W-G. 2001. Types of software evolution and software maintenance. *Journal of software maintenance and evolution: Research and Practice* **13**: 3–30.
- Charrad M, Ghazzali N, Boiteau V, Niknafs A. 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software* **61**(6): 1–36.
- Chintakananda A, McIntyre DP. 2014. Market Entry in the Presence of Network Effects: A Real Options Perspective. *Journal of Management* **40**(6): 1535–1557.
- Cioffi-Revilla C. 2010. Computational social science. *Wiley Interdisciplinary Reviews: Computational Statistics* **2**(3): 259–271.
- Cioffi-Revilla C. 2014. *Introduction to Computational Social Science - Principles and Applications*. Springer: London.
- Cipeluch B, Jacob R, Winstanley A, Mooney P. 2010. Comparison of the accuracy of OpenStreetMap for Ireleand with Google Maps and Bing Maps. In *Accuracy 2010 Symposium*. Leicester (UK).
- Clark KB. 1985. The interaction of design hierarchies and market concepts in technological evolution. *Research Policy* **14**(5): 235–251.
- Clark W, Couldry N, MacDonald R, Stephansen HC. 2014. Digital platforms and narrative exchange: Hidden constraints, emerging agency. *New Media & Society*.
- Coast S. 2015. *The Book of OSM*. CreateSpace Independent Publishing Platform.
- Colfer LJ, Baldwin CY. 2016. The mirroring hypothesis: Theory, evidence, and exceptions. *Industrial and Corporate Change* **25**(5): 709–738.
- Connelly BL, Certo ST, Ireland RD, Reutzel CR. 2011. Signaling theory: A review and assessment. *Journal of Management* **37**(1): 39–67.
- Constantinides P, Chiasson MW, Introna LD. 2012. The Ends of Information Systems Research: A Pragmatic Framework. *MIS Quarterly* **36**(1): 1–19.
- Constantinides P, Henfridsson O, Parker G. 2018. Platforms and Infrastructures in the Digital Age. *Information Systems Research (Special Issue)*: 1–20.
- Creswell JW. 2016. Revisiting Mixed Methods and Advancing Scientific Practices. In *The*

Oxford Handbook of Multimethod and Mixed Methods Research Inquiry, Hesse-Biber SN, Johnson RB (eds). Oxford University Press.

- Crowston K. 1997. A Coordination Theory Approach to Organizational Process Design. *Organization Science* **8**(2): 157–175.
- Crowston K, Howison J. 2006. Core and periphery in Free / Libre and Open Source software team communications. *Institute for Software Research* : 1–7.
- Crowston K, Li Q, Wei K, Eseryel UY, Howison J. 2007. Self-organization of teams for free/libre open source software development. *Information and Software Technology* **49**(6): 564–575.
- Crowston K, Rubleske J, Howison J. 2004. Coordination Theory: A Ten-Year Retrospective. In *Human-Computer Interaction in Management Information Systems*, Zhang P, Galleta D (eds). ME Sharpe: 120–138.
- Csárdi G, Nepusz T. 2006. The igraph software package for complex network research. *InterJournal Complex Systems* **1695**: 1–9.
- Dabbish L, Stuart C, Tsay J, Herbsleb J. 2012. Social Coding in GitHub : Transparency and Collaboration in an Open Software Repository. In *CSCW February 11-15*. Seattle, Washington.
- Dahlander L, Frederiksen L. 2012. The Core and Cosmopolitans: A Relational View of Innovation in User Communities. *Organization Science* **23**(4): 988–1007.
- Dahlander L, O'Mahony S. 2011. Progressing to the Center: Coordinating Project Work. *Organization Science* **22**(4): 961–979.
- Daniel S, Stewart K. 2016. Open source project success: Resource access, flow, and integration. *The Journal of Strategic Information Systems* **25**(3): 159–176.
- Davidovic N, Mooney P, Stoimenov L, Minghini M. 2016. Tagging in Volunteered Geographic Information: An Analysis of Tagging Practices for Cities and Urban Regions in OpenStreetMap. *ISPRS International Journal of Geo-Information* **5**(12): 232.
- Debortoli S, Müller O, Junglas I, Vom Brocke J. 2016. Text Mining For Information Systems Researchers: An Annotated Topic Modeling Tutorial. *Communications of the Association for Information Systems* **39**: 110–135.
- Dranove D, Gandal N. 2003. The Dvd-vs.-Divx Standard War: Empirical Evidence of Network Effects and Preannouncement Effects. *Journal of Economics and Management Strategy* **12**(3): 363–386.
- Eaton B, Elaluf-Calderwood S, Sørensen C, Yoo Y. 2015. Distributed Tuning of Boundary Resources: The Case of Apple's iOS Service System. *MIS Quarterly* **39**(1): 217–243.
- Eaton BD. 2012. *The Dynamics of Digital Platform Innovation : Unfolding the Paradox of Control and Generativity in Apple's iOS*. The London School of Economics and Political Science.
- Eck A, Uebernickel F. 2016. Reconstructing Open Source Software Ecosystems: Finding Structure in Digital Traces. In *Proceedings of the 37th International Conference on Information Systems*. Dublin, Ireland.
- Economides N, Katsamakos E. 2006. Two-Sided Competition of Proprietary vs. Open Source Technology Platforms and the Implications for the Software Industry. *Management Science* **52**(7): 1057–1071.
- Eisenmann T, Parker G, Van Alstyne M. 2011. Platform Envelopment. *Strategic Management Journal* **32**: 1270–1285.
- Ethiraj SK. 2007. Allocation of Inventive Effort in Complex Product Systems. *Strategic Management Journal* **28**: 563–584.

- Ethiraj SK, Levinthal D. 2004. Modularity and Innovation in Complex Systems. *Management Science* **50**(2): 159–173.
- Ethiraj SK, Posen HE. 2013. Do Product Architectures Affect Innovation Productivity in Complex Product Ecosystems? In *Collaboration and Competition in Business Ecosystems*, Adner R, Oxley JE, Silverman BS (eds). Emerald Group Publishing: 127–166.
- Evans D, Schmalensee R. 2016. *Matchmakers: The New Economics of Multisided Platforms*. Harvard Business Review Press.
- Evans DS. 2009. How Catalysts Ignite: The Economics of Platform-Based Start-Ups. In *Platforms, Markets and Innovation*, Gawer A (ed). Edward Elgar: Cheltenham: 99–128.
- Evans PC, Gawer A. 2016. *The Rise of the Platform Enterprise A Global Survey*. The Center for Global Enterprise.
- Farjoun M, Ansell C, Boin A. 2015. Pragmatism in Organization Studies: Meeting the Challenges of a Dynamic and Complex World. *Organization Science* **26**(6): 1787–1804.
- Fichman RG, Santos BL Dos, Zheng Z. 2014. Digital Innovation as a Fundamental and Powerful Concept in the Information Systems Curriculum. *MIS Quarterly* **38**(2): 329–353.
- Fielding NG. 2012. Triangulation and Mixed Methods Designs: Data Integration With New Research Technologies. *Journal of Mixed Methods Research* **6**(2): 124–136.
- Fixson SK, Park JK. 2008. The power of integrality: Linkages between product architecture, innovation, and industry structure. *Research Policy* **37**(8): 1296–1316.
- Fleming L, Sorenson O. 2001. Technology as a complex adaptive system: evidence from patent data. *Research Policy* **30**: 1019–1039.
- Förderer J, Kude T, Mithas S, Heinzl A. 2018. Does Platform Owner’s Entry Crowd Out Innovation? Evidence from Google Photos. *Information Systems Research* : 1–18.
- Fox K. 2012, February 18. OpenStreetMap: ‘It’s the Wikipedia of maps’. *The Guardian*.
- Frenken K. 2006. Technological innovation and complexity theory. *Economics of Innovation and New Technology* **15**(2): 137–155.
- Fuentelsaz L, Garrido E, Maicas JP. 2015. A Strategic Approach to Network Value in Network Industries. *Journal of Management* **41**(3): 864–892.
- Gamma E, Vlissides J, Johnson R, Helm R. 1994. *Design patterns: Elements of reusable object-oriented software*. Addison-Wesley.
- Garud R, Kumaraswamy A. 2003. Technological and organizational designs for realizing economies of substitution. In *Managing in the Modular Age: Architectures, Networks, and Organizations*, Garud R, Kumaraswamy A, Langlois N (eds). Blackwell Publishing: Malden, MA, USA: 45–77.
- Gaskin J, Berente N, Lyytinen K, Yoo Y. 2014. Toward Generalizable Sociomaterial Inquiry: A Computational Approach for Zooming In and Out of Sociomaterial Routines. *MIS Quarterly* **38**(3): 849–871.
- Gasson S, Purcelle M. 2018. Legitimate Peripheral Participation in Hybrid FOSS Community Innovation. In *Proceedings of the 51st Hawaii International Conference on System Sciences*: 4554–4563.
- Gawer A. 2014. Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*. Elsevier B.V. **43**(7): 1239–1249.
- Gawer A, Cusumano M. 2013. Industry Platforms and Ecosystem Innovation. *Journal of*

- Product Innovation Management* **31**(3): 417–433.
- Gawer A, Henderson R. 2007. Platform Owner Entry and Innovation in Complementary Markets: Evidence From Intel. *Journal of Economics & Management Strategy* **16**(1): 1–34.
- Germonprez M, Hovorka DS. 2013. Member engagement within digitally enabled social network communities: New methodological considerations. *Information Systems Journal* **23**(6): 525–549.
- Ghazawneh A. 2012. *Towards a Boundary Resources Theory of Software Platforms*. Jönköping University.
- Ghazawneh A, Henfridsson O. 2013. Balancing platform control and external contribution in third-party development: the boundary resources model. *Information Systems Journal* **23**(2): 173–192.
- Ghazawneh A, Henfridsson O. 2015. A paradigmatic analysis of digital application marketplaces. *Journal of Information Technology* **30**(3): 198–208.
- Goldberg A. 2015. In defense of forensic social science. *Big Data & Society* **July-Decem**: 1–3.
- Goldkuhl G. 2004. Meanings of Pragmatism: Ways to conduct information systems research. *International Business* : 17–18.
- Goldkuhl G. 2012. Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems* **21**(2): 135–146.
- Goodchild MF. 2007. Citizens as sensors: The world of volunteered geography. *GeoJournal* **69**(4): 211–221.
- Gray PH, Parise S, Lyer B. 2011. Innovation Impacts of Using Social Bookmarking Systems. *MIS Quarterly* **35**(3): 629–643.
- Grogan TR, Elashoff DA. 2017. A simulation based method for assessing the statistical significance of logistic regression models after common variable selection procedures. *Communications in Statistics: Simulation and Computation*. Taylor & Francis **46**(9): 7180–7193.
- Grover V, Lyytinen K. 2015. New State of Play in Information Systems Research: The Push to the Edges. *MIS Quarterly* **39**(2): 271–296.
- Hannah DP, Eisenhardt KM. 2018. How Firms Navigate Cooperation and Competition in Nascent Ecosystems. *Strategic Management Journal* : 1–30.
- Hanseth O, Lyytinen K. 2010. Design theory for dynamic complexity in information infrastructures: the case of building internet. *Journal of Information Technology* **25**(1): 1–19.
- Hardin J, Hilbe J. 2003. *Generalized Estimation Equations*. Chapman and Hall/CRC: Boca Raton FL.
- Hedman J, Srinivasan N, Lindgren R. 2013. Digital Traces of Information Systems: Sociomateriality Made Researchable. In *Proceedings of the 34th International Conference on Information Systems*. Milan, Italy: 1–17.
- Helmond A. 2015. The Platformization of the Web: Making Web Data Platform Ready. *Social Media + Society* **1**(2): 1–11.
- Henfridsson O, Bygstad B. 2013. The Generative Mechanisms of Digital Infrastructure Evolution. *MIS Quarterly* **37**(3): 907–931.
- 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.
- Henfridsson O, Nandhakumar J, Scarbrough H, Panourgias N. 2018. Recombination in the

- open-ended value landscape of digital innovation. *Information and Organization*. Elsevier **28**(2): 89–100.
- Hevner AR, March ST, Park J, Ram S. 2004. Design Science in Information Systems Research. *MIS Quarterly* **28**(1): 75–105.
- Hilbe JM. 2011. *Negative Binomial Regression*, 2nd ed. Cambridge University Press: New York.
- Hilbe JM. 2014. *Modeling Count Data*. Cambridge University Press: New York.
- Ho SY, Rai A. 2017. Continued Voluntary Participation Intention in Firm- Participating Open Source Software Projects. *Information Systems Research* **28**(3): 603–625.
- Ver Hoef JM, Boveng PL. 2007. Quasi-poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology* **88**(11): 2766–2772.
- Hosmer D, Lemeshow S, Sturdivant RX. 2013. Model-Building Strategies and Methods for Logistic Regression. In *Applied Logistic Regression*, 3rd ed. John Wiley & Sons: 89–151.
- Howison J, Crowston K. 2014. Collaboration Through Open Superstition: A Theory of The Open Source Way. *MIS Quarterly* **38**(1): 29–50.
- Howison J, Wiggins A, Crowston K. 2011. Validity Issues in the Use of Social Network Analysis with Digital Trace Data. *Journal of the Association for Information Systems* **12**(12): 767–797.
- Huang S *et al.* 2014. Identifying Node Role in Social Network Based on Multiple Indicators. *PLoS ONE* **9**(8): e103733.
- Hukal P. 2017. On the Role Module Interdependencies in Platform Evolution. In *Proceedings of the 38th International Conference on Information Systems*. Seoul, Korea.
- Hukal P, Eck A. 2016. Adapting new capabilities or enhancing functionality? Two sequence patterns of capability redeem in platform ecosystems. In *The 8th International Workshop on Software Ecosystems co-located with the International Conference on Information Systems, Dec 10th 2016*. Dublin, Ireland: 1–24.
- Hukal P, Henfridsson O. 2017. Digital Innovation – A Definition and Integrated Perspective. In *The Routledge Companion to Management Information Systems*, Galliers RD, Stein M-K (eds). Routledge: 360–369.
- Jacobides MG, Cennamo C, Gawer A. 2018. Towards a Theory of Ecosystems. *Strategic Management Journal*.
- Johnson SL, Safadi H, Faraj S. 2015. The emergence of online community Leadership. *Information Systems Research* **26**(1): 165–187.
- Kallinikos J. 2013. The Allure of Big Data. *Mercury Winter*(3): 40–43.
- Kallinikos J, Aaltonen A, Marton A. 2013a. *Information Generativity and Digital Innovation*. Research Report 2013. London, UK.
- Kallinikos J, Aaltonen A, Marton A. 2013b. The Ambivalent Ontology of Digital Artifacts. *MIS Quarterly* **37**(2): 357–370.
- Kallinikos J, Constantiou ID. 2015. Big data revisited: A rejoinder. *Journal of Information Technology*. Nature Publishing Group **30**(1): 70–74.
- Kapoor R, Agarwal S. 2017. Sustaining Superior Performance in Business Ecosystems: Evidence from Application Software Developers in the iOS and Android Smartphone Ecosystems. *Organization Science* **28**(3): 531–551.
- Karhu K, Gustafsson R, Lyytinen K. 2018. Exploiting and Defending Open Digital Platforms with Boundary Resources: Android’s Five Platform Forks. *Information*

- Systems Research* (special issue): 1–19.
- Kauffman S, Macready W. 1995. Technological Evolution and Adaptive Organizations. *Complexity* : 26–43.
- Kazan E, Tan CW, Lim ETK, Sørensen C, Damsgaard J. 2018. Disentangling Digital Platform Competition: The Case of UK Mobile Payment Platforms. *Journal of Management Information Systems* **35**(1): 180–219.
- Kelle U. 2015. Mixed Methods and the Problems of Theory Building and Theory Testing in the Social Sciences. *Oxford Handbook of Multimethod and Mixed Methods Research Inquiry* (November): 594–606.
- Ketokivi M, Mantere S. 2010. Two Strategies for Inductive Reasoning in Organizational Research. *Academy of Management Review* **35**(2): 315–333.
- Kitchin R. 2014a. Conceptualising Data. In *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Sage Publications: London: 1–27.
- Kitchin R. 2014b. The Reframing of Science, Social Science and Humanities Research. In *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Sage Publications: London: 128–149.
- Kuhn T. 2012. *The Structure of Scientific Revolutions: 50th Anniversary Edition*, 4th ed. University of Chicago Press: Chicago/London.
- Lambrecht A *et al.* 2014. How do firms make money selling digital goods online? *Marketing Letters* **25**(3): 331–341.
- Langlois RN. 2002. Modularity in technology and organization. *Journal of Economic Behavior & Organization* **49**(1): 19–37.
- Langlois RN. 2007. The entrepreneurial theory of the firm and the theory of the entrepreneurial firm. *Journal of Management Studies* **44**(7): 1107–1124.
- Lazer D *et al.* 2009. Computational Social Science. *Science* **323**(February): 721–723.
- Lazer D, Radford J. 2017. Data ex Machina: Introduction to Big Data. *Annual Review of Sociology* **43**(1): 1–21.
- Lee J, Berente N. 2012. Digital Innovation and the Division of Innovative Labor: Digital Controls in the Automotive Industry. *Organization Science* **23**(5): 1428–1447.
- Lessig L. 1999. Limits in Open Code: Regulatory Standards and the Future of the Net. *The Berkeley Technology Law Journal* **14**(2): 759–769.
- Lindberg A, Berente N, Gaskin J, Lyytinen K. 2016. Coordinating Interdependencies in Online Communities: A Study of an Open Source Software Project. *Information Systems Research* **27**(4): 751–772.
- Lorino P. 2018. *Pragmatism and Organization Studies*. Oxford Scholarship Online.
- Lusch RF, Nambisan S. 2015. Service Innovation: A Service-Dominant Logic Perspective. *MIS Quarterly* **39**(1): 155–176.
- Lyytinen K. 1985. Implications of Theories of Language for Information Systems. *MIS Quarterly* **9**(1): 61–74.
- Lyytinen K, Rose GM. 2003. The Disruptive Nature of Information Technology Innovations: The Case of Internet Computing in System Development Organizations. *MIS Quarterly* **27**(4): 557–596.
- Lyytinen K, Sørensen C, Tilson D. 2018. Generativity in Digital Infrastructures: A Research Note. In *The Routledge Companion to Management Information Systems*, Galliers RD, Stein M-K (eds).
- Lyytinen K, Yoo Y, Boland RJ. 2016. Digital product innovation within four classes of

- innovation networks. *Information Systems Journal* : 47–75.
- Malone TW. 1988. What is Coordination Theory ? *Technology* : 1–32.
- Malone TW, Crowston K. 1991. Toward an interdisciplinary theory of coordination. *ACM Computing Surveys* **120**(120): 1–45.
- Marres N. 2017. *Digital Sociology: The Reinvention of Social Research*. Polity.
- McIntyre DP, Srinivasan A. 2017. Networks, Platforms, And Strategy: Emerging Views and Next Steps. *Strategic Management Journal* **38**: 141–160.
- Merali Y. 2006. Complexity and information systems: The emergent domain. *Journal of Information Technology* **21**(4): 216–228.
- Miller JH, Page SE. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press: Princeton and Oxford.
- Mooney P, Corcoran P. 2012. The annotation process in OpenStreetMap. *Transactions in GIS* **16**(4): 561–579.
- Mooney P, Corcoran P. 2014. Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors. *Transactions in GIS* **18**(5): 633–659.
- Morgan DL. 2007. Paradigms Lost and Pragmatism Regained Methodological - Methodological Implications of Combining Qualitative and Quantitative Methods. *Journal of Mixed Methods Research* **1**: 48–76.
- Morozoff EP. 2010. Using a line-of-code Metric to understand software rework. *IEEE Software* **27**(1): 72–77.
- Müller O, Junglas I, Brocke J vom, Debortoli S. 2016. Utilizing big data analytics for information systems research: challenges, promises and guidelines. *European Journal of Information Systems* **25**(4): 289–302.
- Murmann JP, Frenken K. 2006. Toward a systematic framework for research on dominant designs, technological innovations, and industrial change. *Research Policy* **35**(7): 925–952.
- Nagaraj A. 2017. Does Copyright Affect Reuse? Evidence from Google Books and Wikipedia. *Management Science* : 1–17.
- Nambisan S, Lyytinen K, Majchrzak A, Song M. 2017. Digital Innovation Management: Reinventing Innovation Management Research in a Digital World. *MIS Quarterly* **41**(1): 223–238.
- Ng I, Smith L. 2012. An integrative framework of value. *Review of Marketing Research* **9**(Special Issue-Toward a Better Understanding of the Role of Value in Markets and Marketing): 207–243.
- Niculescu MF, Wu DJ, Xu L. 2018. Strategic Intellectual Property Sharing: Competition on an Open Technology Platform Under Network Effects. *Information Systems Research* (2018): 1–22.
- O'Halloran KL, Tan S, Pham D-S, Bateman J, Vande Moere A. 2016. A Digital Mixed Methods Research Design: Integrating Multimodal Analysis With Data Mining and Information Visualization for Big Data Analytics. *Journal of Mixed Methods Research*.
- Olleros X. 2008. The Lean Core in Digital Platforms. *Technovation* **28**(5): 266–276.
- Orlikowski WJ, Iacono CS. 2001. Research Commentary: Desperately Seeking the 'IT' in IT Research--A Call to Theorizing the IT Artifact. *Information Systems Research* **12**(2): 121–134.
- Paavola S. 2015. Deweyan Approaches to Abduction? In *Action, Belief and Inquiry—Pragmatist Perspectives on Science, Society and Religion*: 230–249.

- Parker G, Van Alstyne M. 2017. Innovation, Openness, and Platform Control. *Management Science* : 1–18.
- Parker G, Van Alstyne M, Choudary S. 2016. *Platform Revolution: How Networked Markets Are Transforming the Economy - and How to Make Them Work for You*, 1st Edition. W.W. Norton & Co.
- Parker G, Van Alstyne M, Jiang X. 2017. Platform Ecosystems: How Developers Invert The Firm. *MIS Quarterly* **41**(1): 255–266.
- Parker G, Van Alstyne MW. 2005. Two-Sided Network Effects : A Theory of Information Product Design. *Management Science* **51**(10): 1494–1504.
- Parsons E. 2013. The Map of the Future May Not Be a Map! *The Cartographic Journal* **50**(2): 182–186.
- Pentland BT, Feldman MS. 2007. Narrative Networks: Patterns of Technology and Organization. *Organization science* **18**(5): 781–795.
- Phelps C, Heidl R, Wadhwa A. 2012. Knowledge, Networks, and Knowledge Networks. *Journal of Management* **38**(4): 1115–1166.
- Plantin J-C, Lagoze C, Edwards PN. 2018. Re-integrating scholarly infrastructure: The ambiguous role of data sharing platforms. *Big Data & Society* **5**(1): 1–14.
- Porter ME. 1980. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. Free Press: New York.
- Rai A. 2016. Synergies Between Big Data and Theory. *MIS Quarterly* **40**(2): iii–ix.
- Ramm F. 2015. OpenStreetMap Data in Layered GIS Format. GeoFabrik GmbH.
- Ramm F, Topf J, Chilton S. 2010. *OpenStreetMap Using and Enhancing the Free Map of the World*. UIT Cambridge (ed). Cambridge, UK.
- Rapley T. 2015. Sampling Strategies in Qualitative Research. In *The SAGE Handbook of Qualitative Data Analysis*: 49–63.
- de Reuver M, Sørensen C, Basole RC. 2018. The Digital Platform - A Research Agenda. *Journal of Information Technology* : 1–12.
- Rietveld J, Eggers JP. 2018. Demand Heterogeneity in Platform Markets: Implications for Complementors. *Organization Science* (29(2)): 304–322.
- Rochet J-C, Tirole J. 2003. Platform Competition in Two-Sided Markets. *Journal of the European Economic Association* **1**(4): 990–1029.
- Rolland KH, Mathiassen L, Rai A. 2018. Managing Digital Platforms in User Organizations: The Interactions Between Digital Options and Digital Debt. *Information Systems Research* : 1–25.
- Salganik M. 2017. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.
- Sanchez R, Mahoney JT. 1996. Modularity, Flexibility, and Knowledge Management in Product and Organization Design. *Strategic Management Journal* **17**: 63–76.
- Sarker S, Ahuja M, Sarker S. 2018. Work – Life Conflict of Globally Distributed Software Development Personnel : An Empirical Investigation Using Border Theory. *Information Systems Research* **29**(1): 103–126.
- Saxena A *et al.* 2017. A review of clustering techniques and developments. *Neurocomputing*. Elsevier B.V. **267**: 664–681.
- Schilling MA. 2000. Toward a General Modular Systems Theory and Its Application to Interfirm Product Modularity. *Academy of Management Review* **25**(2): 312–334.
- Schreier M. 2014. Qualitative Content Analysis. In *The SAGE Handbook of Qualitative*

Data Analysis: 1–29.

- Schumpeter JA. 1983. *The Theory of Economic Development*. Transaction Publishers: New Brunswick / London.
- Setia P, Rajagopalan B, Sambamurthy V, Calantone R. 2012. How peripheral developers contribute to open-source software development. *Information Systems Research* **23**(1): 144–163.
- Shaikh M, Vaast E. 2016. Folding and Unfolding: Balancing Openness and Transparency in Open Source Communities. *Information Systems Research* **27**(4): 813–833.
- Shapiro C, Varian HR. 1998. Versioning: The smart way to sell information. *Harvard Business Review* (Nov/ Dec): 106–114.
- Simon H. 1962. The Architecture of Complexity. *Proceedings of the American Philosophical Society* **106**(6): 467–482.
- Simon H. 1996. *Sciences of the Artificial*. MIT Press.
- Simon H. 2000. Near-decomposability and the Speed of Evolution. *Industrial and Corporate Change* **11**(3): 587–599.
- Spagnoletti P, Resca A, Lee G. 2015. A design theory for digital platforms supporting online communities: A multiple case study. *Journal of Information Technology* **30**(4): 364–380.
- Svahn F, Mathiassen L, Lindgren R. 2017. Embracing Digital Innovation in Incumbent firms: How Volvo Cars Managed Competing Concerns. *MIS Quarterly* **41**(1): 239–253.
- Tanriverdi H, Rai A, Venkatraman N. 2010. Reframing the dominant quests of information systems strategy research for complex adaptive business systems. *Information Systems Research* **21**(4): 822–834.
- Terlaak A, King AA. 2006. The effect of certification with the ISO 9000 Quality Management Standard: A signaling approach. *Journal of Economic Behavior and Organization* **60**(4): 579–602.
- Tilson D, Lyytinen K, Sørensen C. 2010. Research Commentary —Digital Infrastructures: The Missing IS Research Agenda. *Information Systems Research* **21**(4): 748–759.
- Tiwana A. n.d. Platform Synergy: Architectural Origins and Competitive Consequences. *Information Systems Research* (forthcoming).
- Tiwana A. 2008. Does Technological Modularity Substitute For Control? A Study of Alliance Performance in Software Outsourcing. *Strategic Management Journal* **29**: 769–780.
- Tiwana A. 2014. *Platform Ecosystems: Aligning Architecture, Governance, and Strategy*. Morgan Kaufmann: Amsterdam.
- Tiwana A. 2015. Evolutionary Competition in Platform Ecosystems. *Information Systems Research* **26**(2): 266–281.
- Tiwana A, Konsynski B, Bush AA. 2010. Research Commentary —Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics. *Information Systems Research* **21**(4): 675–687.
- Tsay J, Dabbish L, Herbsleb J. 2014. Influence of social and technical factors for evaluating contribution in GitHub. *Proceedings of the 36th International Conference on Software Engineering - ICSE 2014* : 356–366.
- Ulrich K. 1995. The role of product architecture in the manufacturing firm. *Research Policy* **24**: 419–440.
- Um S, Yoo Y. 2016. The Co-Evolution of Digital Ecosystems. In *37th International*

- Conference on Information Systems Research*. Dublin, Ireland: 1–15.
- Um SY, Yoo Y, Wattal S, Kulathinal RJ, Zhang B. 2013. The Architecture of Generativity in a Digital Ecosystem. In *Proceedings of the International Conference on Information Systems (ICIS)*. Milan, Italy.
- Vaast E, Safadi H, Lapointe L, Negoita B. 2017. Social Media Affordances for Connective Action: An Examination of Microblogging Use During the Gulf of Mexico Oil Spill *41*(4): 1179–1205.
- Vaast E, Walsham G. 2013. Grounded theorizing for electronically mediated social contexts. *European Journal of Information Systems*. Nature Publishing Group *22*(1): 9–25.
- Vandecasteele A, Devillers R. 2015. Improving Volunteered Geographic Information Quality Using a Tag Recommender System: The Case of OpenStreetMap. In *OpenStreetMap in GIScience*, Arsanjani JJ, Zipf A, Mooney P, Heiblich M (eds). Springer Science+Business: Heidelberg/New York: 59–80.
- Varian HR. 2014. Beyond Big Data. *Business Economics*. Nature Publishing Group *49*(1): 27–31.
- Varian HR, Farrel J, Shapiro C. 2004. *The Economics of Information Technology*. Cambridge University Press: New York.
- Van de Ven A, Polley D, Garud R, Venkatesh V. 1999. *The Innovation Journey*. Oxford University Press: New York.
- Venkatesh V, Brown S a, Bala H. 2013. Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly* *37*(1): 21–54.
- Venkatesh V, Brown S, Sullivan Y. 2016. Guidelines for Conducting Mixed-methods Research: An Extension and Illustration. *Journal of the Association for Information Systems* *17*(7): 435–495.
- Wareham J, Fox PB, Cano Giner JL. 2014. Technology Ecosystem Governance. *Organization Science* *25*(4): 1195–1215.
- West J. 2003. How open is open enough? Melding proprietary and open source platform strategies. *Research Policy* *32*(7): 1259–1285.
- Wicks AC, Freeman RE. 1998. Organization Studies and the New Pragmatism: Positivism, Anti-positivism, and the Search for Ethics. *Organization Science* *9*(2): 123–140.
- Winkelmann R. 2008. *Econometric Analysis of Count Data*. Springer: Berlin/Heidelberg.
- Woodard CJ, Ramasubbu N, Tschang FT, Samabamurthy V. 2013. Design Capital and Design Moves: The Logic of Digital Business Strategy. *MIS Quarterly* *37*(2): 537–564.
- Wooldrige J. 2016. *Introductory to Econometrics - A Modern Approach*, 6th Editio. Cengage Learning: Boston, MA.
- Yin RK. 2015. Causality, Generalizability, and the Future of Mixed Methods Research. In *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry*, Hesse-Biber SN, Johnson RB (eds). Oxford University Press: 1–25.
- Yoo Y. 2010. Computing In Everyday Life: A Call For Research on Experiential Computing. *MIS Quarterly* *34*(2): 213–231.
- Yoo Y. 2013. The Tables Have Turned: How Can the Information Systems Field Contribute to Technology and Innovation Management Research. *Journal of the Association for Information Systems* *14*(5): 227–236.
- Yoo Y. 2015. Design in the Generative Economy. *Research-Technology Management* *58*(2): 13–20.
- Yoo Y, Boland RJ, Lyytinen K, Majchrzak A. 2012. Organizing for Innovation in the

- Digitized World. *Organization Science* **23**(5): 1398–1408.
- Yoo Y, Henfridsson O, Lyytinen K. 2010. The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research* **21**(4): 724–735.
- Zachariadis M, Scott S, Barrett M. 2013. Methodological Implications of Critical Realism for Mixed-Methods Research. *MIS Quarterly* **37**(3): 855–879.
- Zeng X, Wei L. 2013. Social ties and user content generation: Evidence from Flickr. *Information Systems Research* **24**(1): 71–87.
- Zhang Z, Yoo Y, Kulathinal R, Wattal S. 2015. *Characterizing Generative Digital Artifacts: The Case of Web Mashups*. Organizational Genetics Working Paper. Philadelphia, PA, USA.
- Zhang Z, Yoo Y, Wattal S, Zhang B, Kulathinal R. 2014. Generative Diffusion of Innovations and Knowledge Networks in Open Source Projects. In *Proceedings of the International Conference on Information Systems (ICIS)*. Auckland.
- Zittrain J. 2006. The Generative Internet. *Harvard Business Review* **119**(7): 1974–2040.
- Zittrain J. 2008. *The Future of the Internet - And How to Stop It*. Yale University Press: New Haven.