

**THESIS SUBMITTED IN PARTIAL FULFILMENT FOR THE
DEGREE OF D.PHIL. IN MANAGEMENT RESEARCH**



**PRODUCT PUBLICS AND THE EARLY MOMENTS OF NASCENT
MARKETS**

Timothy R. Hannigan
Green Templeton College
Saïd Business School, University of Oxford
Trinity Term 2012

PRODUCT PUBLICS AND THE EARLY MOMENTS OF NASCENT MARKETS

Timothy R. Hannigan, Green Templeton College, University of Oxford
Submitted in partial fulfilment of the requirements for the degree of D.Phil. in
Management Research
Trinity Term 2012

Abstract

Product markets have been found to be socially constructed as industry logics, product ontologies and organising visions. Information intermediaries play an important field-level role in determining the categories and ontologies of emerging products. Within shared cognitive structures, sellers and buyers come together in mutual understandings on how products are used, what features are important and what counts as value.

Although market information is often legitimated through objective scientific methodologies, it can also appear as uncertain knowledge, or rumour. Contrary to the often seen conception of rumour as crowd-like behavioural contagion, it is presented here as an informal message lacking a reliable sponsor that is developed in an uncertain situation, but still may be subject to critical reasoning.

This thesis investigates the ambiguity around the fuzzy front-end of product markets, as social media and blogs that trade in rumour and proposition before any official announcement by firms. Through a process of improvised news, these online collectivities, as product publics, deliberate and collectively generate expectations as meaning structures. By tracking discourse from tech blogs, I trace the emergence of meaning in the tablet device product ontology. This occurred well before product introductions by firms who eventually came to form the market.

Keywords: *new product development, product ontologies, markets as fields, network text analysis, concept networks, early moments, nascent markets, computational linguistics, rumour, sensemaking*

Acknowledgements

It is a wonderful thing to be writing acknowledgements as the final step closing out a lengthy project. It is this moment when I look back and reflect on collective support and effort which enabled this work to come to fruition. A dissertation is indeed a project, but it is also an artefact of a process. And at some point along the marathon, the process called "education" ceased to be about following the preordained knowledge-base. It became the process of learning to become a scholar, entailing not only considering the socialisation of scholarship, but of the commitment, dedication and ethics that govern the collective endeavour. As I come to the end of my Doctorate, I can report with some confidence that the world of scholarship is fundamentally based on the community of support. On reflection this confirms my belief that each and every academic career is made possible only through the support, mentoring and guidance that go far beyond the scope of set roles. It goes without saying that engaging in such a line of work is daunting.

Along my journey, there were many people that I would like to thank for generously helping in innumerable ways. I had the pleasure of having two supervisors whilst at Oxford, Dr Victor Seidel and Dr Marc Ventresca. Victor graciously invested much energy in seeing the value of my project early on. I am grateful for his tenacity and encouragement in helping to show me a path and ensure I had a direction. Working through pitfalls and unexpected roadblocks of data collection are greatly helped by a supportive supervisor. To Dr Marc Ventresca, I am grateful for his mentorship, guidance and an unconditional encouragement. He generously supported me from the very beginning, interviewing at Oxford, onward to working through the dissertation and finally seeing it through. His intellectual guidance has provided me a platform from which to explore unconventional ties between Sociology, Innovation and Strategy. I continue to value his council, support, and creativity in entering an academic career. Other faculty members at Said Business School also provided me guidance and support. I particularly want to thank Felix Reed-Tsochas/CabDyn group, Richard Whittington, and Thomas Noe.

My journey was also marked by a memorable and critically insightful term visiting Northwestern University in Chicago. It was there that I really explored the method of Network Text Analysis and found strong ties to publics and cultural sociology. This trip was made possible through the generosity of Noshir Contractor, SONIC, Gary Alan Fine, Klaus Weber, Paul Leonardi, and the Sociology department. As well, friends Lorne Kligerman and Kate McGregor provided a great level of support as friends opening their homes and helping to set us up in a foreign city.

To my Oxford friends: Alex Budzier, George Chondrakis, Masahiro Kotosaka, Benedikt Langner, Jens Krause, Tomas Farchi, Tarek Cheniti, Charles Laurie, Emanuele Ferragina and Roy Nyberg. We travelled these roads together and it is wondrous to see a dynamic of collegiality outweigh any sense of competition or derision. It's not easy trying to persuade the upside of rumour. These friends

supported me with an open mind and sense of curiosity. My long-time friends in Canada were always there to encourage, and I value their commitment greatly. Rizwan Jiwan played a critical role in supporting several iterations of this project. Matthew Hellin, Matthew Shogilev and Kyle Waldman were always willing to explore ideas over a good cup of coffee. To Johannes Tulusan, Ferras Ng and Suyog Peshkar, I thank you for your encouragement all the way through the process, beginning from LSE. To Kenneth and Claribelle Rhode, and Charles Wahab, I am grateful for your support and encouragement for many years, and particularly in helping to make it over the final moments.

To Edilma and Gerardo Jiménez, I am grateful for your support, encouragement and providing a comfortable working environment. Writing is certainly enriched by tasty arepas and Colombian coffee.

I am eternally grateful to my family for their affection and support all the way through my DPhil and graduate education. My brother TJ is a strategic management expert and has always been a pillar of support. I thank my sisters Maeve and Olivia for their never-ending encouragement and help with editing. They have both always sustained a wonderful belief in my ability to see things through, no matter how difficult it may seem to me.

To my parents Ruth and John, I am amazed by your unwavering support and encouragement. From a general level of my interest in graduate education to the inception of the thesis idea seven years before I began, I draw tremendous inspiration from you both. Your passion for scholarship and the world of ideas has been incredible to me. On an implementation level, the number of drafts you edited late in to the night were innumerable, as were the many hours spent working through the nascent ideas. I thank you for your extraordinary commitment and always being there.

Finally, to Viviana Jiménez, my co-pilot in this journey. Your support, affection and encouragement was nothing short of magnificent. From talking through concepts for hours on end, to coding and transcribing data, and editing chapter upon chapter, your involvement was integral throughout every page of this document. You are a talented editor and have a wonderful capacity to explore ideas and places. Your sacrifices and dedication to seeing this through are noted and should serve inspiration to the rest of us. My gratitude to you in this of the utmost height. You demonstrate single-handedly how a Doctorate is only made possible through the help of others.

List of Contents

Abstract	ii
Acknowledgements	iii
List of Contents	v
List of Tables	ix
List of Figures	x
List of Abbreviations	xii
Chapter 1: Introduction	1
1.1 Introduction	1
Chapter 2: Theoretical introduction & literature review	9
2.1 Introduction	9
2.2 New product development	12
2.3 User innovation	14
2.4 Information flows and innovation	16
2.5 Interorganisational perspective on product markets	25
2.5.1 Shared knowledge structures amongst market actors	28
2.4.2 Social construction of technology	31
2.4.3 The market as an organisational field	32
2.4.4 Market information	34
2.4.5 Early market information as rumour	37
2.4.6 How product markets emerge as nascent markets	41
2.4.7 When product ontologies emerge	43
2.5 Towards a discursive view of product market emergence	46
2.5.1 Organizing vision	47
2.5.2 Product market as discourse and product ontology	48
2.5.3 A sociological perspective on publics	52
2.5.4 Studying the emergence of a product ontology in a product-public	55
2.6 Finding product publics online	57

2.7 Studying the emergence of a product ontology	62
Chapter 3 - Methodology	66
3.1 Introduction	66
3.2 Studying early moments	69
3.2.1 Studying early moments of a product market	69
3.2.2 Studying product publics	70
3.2.2 Studying emerging product ontologies as relational structures	76
3.3 The tablet computer and tech blogs	80
3.3.1 The tablet computer	80
3.3.2 Tech blogs	83
3.4 Data collection	87
3.4.1 Introduction	87
3.4.2 Sampling a web forum to find a public: Macrumors.com	87
3.4.3 Sampling from a product public around a tablet: tech blogs	90
3.4.4 Projecting concept networks from text	92
3.5 Developing a novel method for projecting concept networks	96
3.5.1 Introduction	96
3.5.2 Concept networks	96
3.5.3 Constructing concept networks	104
3.5.4 Simple example of a concept network	107
3.6.5 Natural language processing workflow: POS tagging & chunking	114
3.5.6 Advanced rules for method in collapsing and labelling nodes from noun phrases	123
3.5.7 Product ontologies as concept networks	124
3.5.8 Product ontology as sets of cognitive classes	125
3.6 Data analysis	127
3.6.1 Introduction	127
3.6.2 Analysing and measuring structural dynamics in a user forum	128
3.6.2.1 Analysing proportions of activity across thread categories	130
3.6.2.2 User overlap	131
3.6.2.3 Timing dynamics	132

3.6.2.3 Measuring public and community activity	132
3.6.3 Analysing and measuring concept networks	135
3.6.3.1 Modularity as a measure of discursive cohesiveness	137
3.6.3.2 Analysis of cognitive class centrality in longitudinal study	142
3.7 Applications of novel method	145
3.7.1: Introduction	145
3.7.2: Demonstrating propositions in an analysis	145
3.7.3: Sample analysis of concept networks from corpus	147
3.7.4: Sample results	151
3.8 Conclusion	154
 Chapter 4 - The ecology of a tech blog: communities and product publics	 157
4.1 Introduction	157
4.2 Descriptive measures	159
4.3 Proportions of activity	165
4.3.1 Thread categories: aggregates	169
4.4 User Overlap	173
4.5 Timing dynamics of aggregate thread categories	178
4.5.1 Life-cycles of threads	181
4.6 Analysis	183
4.7 Discussion	186
 Chapter 5: The emergence of product ontologies	 188
5.1 Introduction	188
5.3 Results	190
5.3.1 Concept network projections as longitudinal data	190
5.3.2 Coding for cognitive classes	206
 Chapter 6: Discussion & Conclusion	 211
6.1 Introduction	211
6.2 Implications and Contributions	214
6.2.1 Product Ontology	216
6.2.2 Product Publics	220

6.2.3 Community	223
6.2.4 Online communities and publics	224
6.2.5 Method: Noun phrases in concept network analysis	228
6.2.6 Method: Considering rumour and speculation as data	228
6.2.7 Contributions to innovation studies	230
6.3 Limitations	233
6.4 Conclusions	234
References	236
Appendices	267
Annexe 1: Data collection methods	267
Annexe 2: Alphabetical list of part-of-speech tags used in the Penn Treebank Project	269
Annexe 3: Advanced rules for method in collapsing and labelling nodes	270
Annexe 4: Descriptions of tech-blogs from an aggregator	271

List of Tables

Table 3.1: Information crawled from the Macrumors.com Jan 1 2007-Jan 31, 2008	89
Table 3.2: Tech Blog descriptions	91
Table 3.3: Thread Categories on macrumors.com.	129
Table 3.4: Sample of Cognitive Classes coding - 2009 Q3	144
Table 3.5: Descriptive statistics of base concept networks	147
Table 3.6: Descriptive statistics of sample and combined concept networks	151
Table 3.7: Effect of modularity from variance of sampling conditions	152
Table 4.1: Description of overall activity	159
Table 4.2: Thread categories and thread frequencies, total users, posts	162
Table 4.3: Correlations between variables frequencies in thread categories	164
Table 4.4: Post, Thread and User activity in top Thread sub-categories	164
Table 4.5: Comparing two thread sub-categories	169
Table 4.6: Aggregate thread categories: thread and posting activity	170
Table 4.7: Measuring skew and kurtosis of thread-lifespans across thread categories	181
Table 6.1: Conceptually separating Community & Public	226

List of Figures

Figure 2.1: Literature Review flow chart	10
Figure 3.1: Full concept network for Sample Tablet Rumour	108
Figure 3.2: Natural Language Processing workflow	110
Figure 3.3: Parse tree for first sentence of sample tablet rumour	113
Figure 3.4: Parse tree for second sentence of sample tablet rumour	114
Figure 3.5: Concept network for Blurb Sentence 1	119
Figure 3.6: Concept network for Blurb Sentence 2	119
Figure 3.7: Concept network for Sample blurb	121
Figure 3.8: Spectrum of mechanisms explaining Users propensity to post	133
Figure 3.9: Modularity explained (source: Shwed & Bearman, 2010)	139
Figure 3.10: Sample article on apple tablet rumour, Q42009	149
Figure 3.11: Sample article on Dell tablet PC rumour, Q42007	149
Figure 3.12: Sample article on apple tablet rumour, Q32009	150
Figure 3.13: Sample article on Android tablet rumour, Q32009	150
Figure 4.1a: Histogram of the number of Posts per Thread	160
Figure 4.1b: Cumulative Distribution Function (CDF) of the Number of Posts per Thread	160
Figure 4.2a: Histogram of the number of Posts per User	161
Figure 4.2b: Cumulative Distribution Function (CDF) of the Number of Posts per User	161
Figure 4.3: Normalised percentages of thread sub-categories, Post counts and Unique User participation	166
Figure 4.4: Sub-categories containing largest percentage of posts	168
Figure 4.5: Thread and Post counts across Thread Categories	171
Figure 4.6: User activity overlap across thread categories	174
Figure 4.7: Dendrogram showing user activity overlap in thread categories	175
Figure 4.8: Time of postings in all threads (initial post at time = 0)	178
Figure 4.9: Box plots of post timing across thread categories: as year, month, week	180
Figure 4.10: All threads with > 200 posts; lifespans as posts per day	181
Figure 5.1: Posts containing keyword *tablet	190
Figure 5.2 - Concept network from "Tablet" Sample, 2008-Q4	192
Figure 5.3: Concept network from "Tablet" Sample, 2008-Q4, filtered nodes, degree > 19	194
Figure 5.4: Concept network from "Tablet" Sample, 2009-Q4	196

Figure 5.5 - Concept network from “Tablet” Sample, 2009-Q4, filtered based on degree > 39	197
Figure 5.6 - Concept network from “Tablet” Sample, 2010-Q4	198
Figure 5.7: Concept network from “Tablet” Sample, 2010-Q4, filtering based on degree > 96	200
Figure 5.8: Concept network from “Tablet” Sample, 2010-Q4, filtering based on degree > 217	201
Figure 5.9: Modularity Scores For Overall Concept network from “Tablet” Sample & Network Evolution	203
Figure 5.10: Distribution of Cognitive classes in a product ontology	207
Figure 5.11: Distribution of Cognitive classes in a product ontology - trends	208

List of Abbreviations

NPD	New Product Development
GUI	Graphic User Interface
OSS	Open Source Software
TF-IDF	term frequency-inverse document frequency
CDF	cumulative density
IDE	Innovation Disclosure Events
CRA	Centering Resonance Analysis
POS	Parts Of Speech
NLP	Natural language processing
STS	Science, Technology and Studies

Chapter 1: Introduction

1.1 Introduction

The emergence of product markets has come to be understood from several perspectives. One view, under the banner of new product development (NPD), considers the efficacy of a firm developing a product (Brown & Eisenhardt, 1995). The primary focus in this view is on optimising critical production processes and controlling information flows inside the firm. The comparative product space is given minor attention as part of either the early stages of the “fuzzy-front end” (Khurana & Rosenthal, 1998), or the downstream marketing function later on in the process. The product that is sold to market can be seen as a rational outcome of the firm’s processes that includes gathering specifications and implementing them. External information to the firm is considered from a distance and in a controlled manner (Brown & Eisenhardt, 1995).

This view is augmented by the early work in user-innovation (von Hippel, 2005), where the 'lead user methodology' outlines how external actors experience needs before the emergence of a mainstream market, and how firms can tap into this as a resource (von Hippel, 1988). This brought external actors into the product development process, including suppliers and users as sources of innovation.

However, members of the firm are still considered to be experts who,

"determine key trends in both the technology and market areas, and then search for 'lead users' who can identify viable design specifications" (Dougherty, 1990).

In this literature, the lead users experience “needs well in advance of the broader market” (von Hippel, 1988), and can potentially be a resource to tap to understand market needs in advance.

A meso level view of product markets comes from the strategy literature, where product markets follow a technology life cycle. This derives from an industrial economics tradition that considers a stylised evolutionary process of variation and selection coming from supply and demand dynamics (Geroski, 2003). This model is simplified as the s-curve (Foster, 1985) where market emergence is depicted as a logistic growth function over time. The first stage is a period of experimentation and ferment, with many different designs being produced. Selection occurs when a particular product design achieves relatively greater sales and other producers modify their offerings to mimic it. This is where the market settles around a dominant design of the product. Once a design is selected through demand signals and economies of scale, product innovations decline and process innovations increase (Utterback & Abernathy, 1975). The dominant design becomes taken-for-granted by producers in the market and subsequent innovations are largely incremental as competition settles around a performance metric (Christensen, 1997). The collection of attributes and dimensions that make up the essential characteristics of the dominant design are part of structured collective understandings and beliefs in the market and industry (Porac et al., 2002). A subsequent new market will emerge when there is a technological discontinuity and new designs emerge that significantly differ from the incumbent dominant design.

The strategy literature has often looked at the stability of industries and boundaries of markets (Porter, 1980; McGahan & Porter, 1997; Santos & Eisenhardt, 2009; Williamson, 1999). One stream within this is a cognitive constructionist view that draws on economic sociology and the social construction of technology to explain

product markets in terms of collective cognitive structures. This view is concerned with consensus around product and social actor definitions as the market settles around the embeddedness of social relations (Fligstein, 1996, 2002; Kennedy, 2005; Porac et al., 2002). The market is a field which brings together disparate groups of actors to consolidate their activities (Anand & Peterson, 2000). This view explicitly considers the shared understandings that constitute the core of the industry. In this, the firm develops products in the shared arena of a broad product market, but settles into market niches. The shared cognitive space of the market thus serves as a platform (Adner, 2012) of belief systems facilitated by market information regimes (Anand & Peterson, 2000). Embedded within this cognitive structure is a market core that provides a common definition of the product in a “collective technology frame” (Kaplan & Tripsas, 2008). Market stratification comes about when firms cluster around frames that partition their shared spaces. The bulk of this work focuses on analysing the stability of these niches, but not to the cycle of emergence, innovation, and change.

In the technology lifecycle work (Utterback & Abernathy, 1975; Anderson & Tushman, 1990; Utterback, 1996), it still remains unclear what characterises discontinuities that shift an industry from a dominant design phase into a new era of ferment. The economic explanation begins with a recognition that disruptive innovations emerge once an industry logic of performance changes (Christensen, 1997), however this shift in logic is taken as given in such analyses. Any theory of market emergence needs to consider how such change occurs, particularly when there is an existing institutional arrangement.

Work in the cognitive constructionist view has begun to expand thinking on discontinuities by considering the link between shifting customer preferences and industry cognitive frames. Tripsas (2008) defines customer preference trajectories as being in parallel with technology trajectories. The claim here is that technology trajectories are coupled with cognitive frames held by aggregates of consumers. Changes in customer preferences are shown to be due to movements in the socio-political environment, shifting interdependence of a modular system, evolution of customers over time through increased size and complexity, and proactive moves by producers to shape preferences through the media. It highlights that customer preference trajectories may be influenced by exogenous shocks from the broader cultural and institutional environment, however, this does not consider how such shifts can be harnessed strategically as sources of innovation.

Theory around the emergence of product markets has largely focused around economic approaches to dominant design. As such, emergence begins when the producers sell to the market. In conceptualising a product market as organisational field (Fligstein, 2002; Anand & Peterson, 2000), emergence can be thought of as occurring earlier along social dimensions. The field construct has several treatments, such as a level of analysis (Scott, 2000: 83) couched between society and an organisational population. It has also been considered by Bourdieu (1977) as a social field that is not restricted to organisations, and instead being understood in relation to a “habitus.” The market as a field is both social and organisational.

One approach to thinking about such a field ties the notion of a “market information regime” (Anand & Peterson, 2000) to organising product categories that enable the

settling of a community of key actors in the market (DiMaggio & Powell, 1983; Fligstein, 2002). This can be seen as similar to Hoffman's (1999) view of an organisational field as a discursive arena, but continues to emphasise the legitimacy of key actors. It allows us to think about the social relations and role orderings of the producer network (White, 1981), as well as the legitimacy that keeps such a cognitive structure intact. Established products and brands matter as cognitive referents for the product definition (Santos & Eisenhardt, 2009). Yet, in thinking about the market as a field, this is clearly a post-hoc view of a stable arrangement. In the nascent state of a market, ambiguity has yet to be resolved (Santos & Eisenhardt, 2009; Alrich & Fiol, 1994; Kennedy, 2008). It is in these early moments, that the product has not been defined, and main producers have yet to form into a community of industry.

Within the cognitive constructionist view, the focus has been on the formation of the producer network. As producers signal to one another with price, they adjust their tactics to explore niches (White, 1981). Recognition of inclusion in the industry network brings legitimacy and market power. In drawing upon legitimacy, this depends on work on structure from institutional theory. For organisational fields, the focus is on the structure and reputation of key actors. The more stable the community of producers becomes, the more the field is understood and institutionalised. Community is the social structure, with key actors making up the dense core, and peripheral actors broadly supporting it.

This focus on the producer network has been critiqued as neglecting other actions of institutional actors such as consumers (Ansari & Phillips, 2011). By focusing on

industry structure and competition, this work could be faulted for ignoring Granovetter's (1985) argument for embeddedness in markets as increasing trust between producers and consumers. This disagreement could have potentially sent the literature back into classic debates of market genesis from supply push or demand pull (Geroski, 2003; Mowery & Rosenberg, 1979), but this was averted by an emphasis on market intermediaries. In the early stages of markets, recognition of the emerging producer network by market intermediaries is important for firm legitimacy (Kennedy, 2008). In stable markets, intermediaries stabilise market interfaces by building an information regime that includes relative orderings such as ranking systems. In this light, product categories become institutions.

White (1981) and Porac et al. (1989) both emphasise social views of a market being constituted through producer defined rivalries. This leads to a cognitive network (Krackhardt, 1987) of legitimate producers. Kennedy (2005, 2008) showed this empirically with a nascent market and Zuckerman (1999) with stable ones. However, there is clearly more to the shared meaning structure than simply a social network of producers as organised by market intermediaries. What cognitive forms are to be embedded in the semantic definition matter. When product attributes are gathered as a constellation around the product definition, then notions of centrality in a product category for a producer are beyond being just mentioned next to key actors. We know that intermediaries often organise the product market by establishing a credible presence with easily interpretable market information framed in a neutral objective manner. But the market information regime also embeds a conceptual system that enables market actors to make sense of the shared arena.

Rosa et al., (1999) documented the emergence of the minivan product category as a public conversation between producers and consumers. In a span of two years, the minivan stabilised as a coherent product category. The conceptual system of meaning in the market shifted so that the hybrid category of car and truck became cognate. This speaks to a shared belief system of the industry that packaged up product attributes, use-cases and buyer characteristics into a taken-for-granted frame. It is a related argument to what is described as dominant design in the technology life cycle (Utterback & Abernathy, 1975; Tushman & Anderson, 1986), however, the definition is not merely a consensus on artefact design and performance metrics.

The product definition question occurs at both the level of the broader industry and the niche product market. In the case of the music industry (Anand & Peterson, 2000), the industry was made up of five major conglomerates as well as a community of smaller producers. Product categories encapsulated musical genres such as *Heavy Rock*, *Country* and *R&B*. Each of those categories contains a cognitive definition that affirms a song or artist's membership within the genre. Performance metrics are defined relative to the cognitive structure of the category. This has been addressed only loosely by the dominant design literature, but it is a critical question in what brings a product market together.

Cognitive structures stabilise in product markets and can occur at many levels. Tripsas (2008) developed arguments about customer preference trajectories being in parallel with technology trajectories and Kaplan & Tripsas (2008) discussed the "collective technology frame" of industries. The work of Porac et al. (2002) was a

focused effort in the same vein that outlined the product ontology as a cognitive structure that incorporates meaning and definition in a product category. This extends thinking about the cognitive space from the claiming and boundary work (Santos & Eisenhardt, 2009), and into a tangible interorganisational knowledge structure (Rosa et al., 1999; Porac et al., 2002). Therefore, it is possible to move from a focus on the politics of a producer network to focus on the meaning of a product ontology. In order to study the product ontology, we need to consider how to operationalise the concept.

The rationale for this thesis comes from considering the early moments of a nascent product market, where neither legitimacy or the community structures are necessarily present. Both of these are prerequisites for stable market information. Anand & Peterson (2000) claim that in some regimes, market information may be found in rumour and speculation. However, this seems to conflict with the claim that market information works through objective framing. I propose that this is overcome by considering the claim that the constitution of fields is information-regime dependent (Anand & Peterson, 2000). In the early moments of a product market, rumours are part of a collective sensemaking process of “improvised news” (Shibutani, 1966). There is an absence of objective information, however, the field will begin to focus common attention on rumours. Despite the risk of the information being incorrect, paying attention to it may be a pragmatic strategy required by involvement in early moments. Although standardised information such as rankings may not be present, attention may still be interlocked by a developing discourse around product ontologies.

Chapter 2: Theoretical introduction & literature review

2.1 Introduction

In this chapter I develop the claim that a product ontology is a knowledge structure that stabilises meaning. Mohr's (1998) work on measuring meaning structures becomes relevant in thinking about the product ontology that envelopes essential elements underpinning a product market. Thus, there is a research gap in tracing product market emergence, but tracing the emergence of meaning structures. Figure 2.1 shows the conceptual flow of the chapter as it moves from views on Product Development and Product Market to a Discursive view of product market emergence.

This chapter traces how innovation and the development of product markets have come to be understood. The first section looks at product innovation, both inside and outside the firm. This begins with a look at a stream known as New Product Development (Brown & Eisenhardt, 1995). Much of this work looks at the critical success factors and performance of processes in individual firms developing successful products. Related work in user-innovation takes the locus of product innovation outside the firm. This describes a paradigm of innovation based on broadcast search (Jeppesen & Lakhani, 2010). Lead-users have keen "local information" that is pertinent to the generation of product concept ideas. However, user innovation also leads to new forms of organisation such as open-source software and user communities. How local information and product concepts are settled upon is a social process. The next section shifts levels and considers interorganisational perspectives on product innovation.

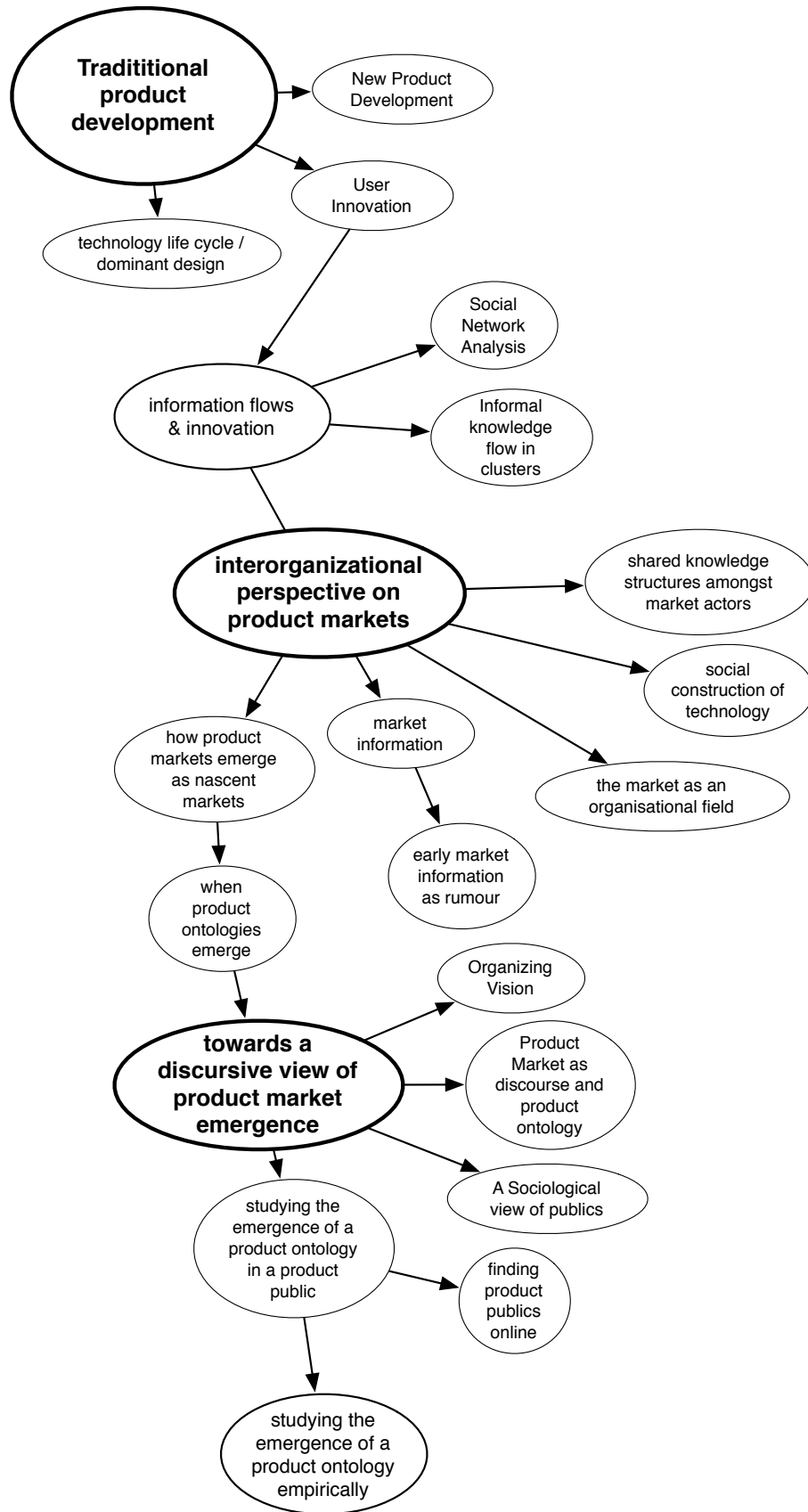


Figure 2.1: Literature Review flow chart

The main works here come from evolutionary economics, organisational fields, market information and product ontologies. The evolutionary economics work is based on a stylised evolutionary process of variation, selection and retention. This helps to explain industry and technology emergence through incentives and signals, but under-theorises social action. Conversely, the remaining other works bring in social structure, institutional theory and organisational cognition. These are held together by the social mechanism of legitimacy and the structure of community.

I build the basis for empirical study by suggesting that to follow the evolution of product ontologies, we should focus on discourse from emergent issue publics (Dewey, 1927; Mische, 2009). Communities are stable forms made up of social relations, and publics are temporary, ephemeral gatherings of actors who suspend their identities and are linked by a common interest. The latter is proposed to be the basis for a new discursive perspective on product market emergence. Finally, I address the research question of the thesis and then I outline the structure of the thesis.

2.2 New product development

The literature on new product development has been studied largely through internal firm processes as critical success factors. Brown & Eisenhardt (1995) and Ernst (2002) conducted wide-ranging reviews of the literature and found that the most common unit of analysis was the product development process. In this organisational-level work, the behaviour of actors is seen to affect successful outcomes. Major themes include information flow, boundary setting and control.

Brown & Eisenhardt (1995) partitioned the literature into themes covering *rational planning*, *the communication web*, and *disciplined problem solving*. All three streams cover the efficacy of firm processes. Determining the market space in which a firm will compete is seen as a factor, but often this space is taken as given. Under the *communication web* theme, the focus is on communication and information flow, and this can include both within the firm and external actors. This work does not explicitly consider the social networks that act as information conduits. Instead, this draws upon information processing theory, where gatekeepers manage the boundary of the firm. Given the limited processing abilities of gatekeepers, information is optimised for amount and some amount of diversity. Under the *rational planning* and *disciplined problem solving* themes, the focus remains on optimising the process. External information is streamlined to knowledge specialists such as suppliers. Therefore, external information flow is considered from a controlled and distanced perspective.

Ambiguity in this literature can be seen as anathema to a successful product development process. The process should rather be well planned and rationally

implemented. However, It is recognised that early on in the product development process, firms often suffer from a fuzzy-front end. This phase has been characterised as ambiguous and can severely hamper the process. Khurana & Rosenthal (1997) suggested this fuzziness is caused by foundational issues surrounding product strategy, portfolio planning, and organisational structure. At the project level, this resulted in fuzziness around the product concept, the product definition, value chain considerations and project planning. The solution to dealing with ambiguity and fuzziness is more focus and control.

Thus, whilst the new product development literature focuses upon internal processes, it does acknowledge that interaction with external actors such as customers and suppliers in the innovation process are a source of potential competitive advantage. Moenardt et al., (2000) studied communication flows with international research and development teams and found that communication with external parties needs to be actively managed, with the risk of developing flawed specifications. This fits into the *communication web* theme that Brown & Eisenhardt (1995) identified, where the independent variable in these studies was always communication. They resolved that frequent task oriented communication with firm outsiders such as customers and suppliers exposed project teams to new information with diverse viewpoints that had a positive affect on innovation. Several studies in the literature emphasise the gains of such external communication (Clark & Fujimoto, 1991; Katz, 1982; Katz & Tushman, 1981), but this is perhaps expressed most in the work on user innovation (von Hippel, 1988).

2.3 User innovation

Much of the NPD literature has focused on technology and internal firm processes, but there has been another stream of literature focused on product development of social actors external to the firm. Under the banner of user innovation, this includes user-toolkits, open source software, and crowdsourcing with innovation contests.

Each of these empirical phenomena contain governance structures, and incentives for participants. The firm is under-theorised in this perspective, in favour of communities that organise economic production.

A central idea in user-innovation is that of the lead-user (von Hippel, 1988). This is a type of user whom will benefit directly from a solution, and experience the need before the general market. This is to say they benefit directly through use, not as a manufacturer seeking pecuniary rewards as a result of manufacturing and sales. Von Hippel (1988) suggests these users can serve as a type of “need-forecasting laboratory” for the market. Users are sources of innovation despite not having direct pecuniary incentives.

The user-innovation literature offers a relevant perspective on how external actors can be important to the innovation process. Here, constructs such as lead-users, free-revealing and user-communities show how these external actors to the firm innovate. When investigating the question of the sources of innovations, this literature found cases of both “user-firms” (von Hippel, 1986, 1988) and the more intuitive “hobbyist-users” (Shah, 2000; Baldwin et al., 2006; Jeppesen & Frederiksen, 2006). The user-firms are those who experience an emergent need for a tool during

production that is not otherwise available and create it for use. Some of these custom-made tools never get marketed as separate products (von Hippel, 1988). The hobbyists, as individual users, experience needs in *real-world use*. This “local information” they derive from such experience is a sticky (von Hippel, 1988) form of tacit knowledge and thus difficult to formalise explicitly (Polanyi, 1967). As a result, traditional market research methods will have significant difficulties extracting these insights. The lead-user method (von Hippel, 1988; Luthje & Herstatt, 2004) was offered as a way to overcome this hurdle and incorporate the optimal type of user into market research by the firm.

Users with similar interests may come together to form communities with the intent of advancing a common activity (Franke & Shah, 2003). Within the community, user-innovators freely reveal their modifications, and receive important advice and assistance from one another (von Hippel, 2001; Jeppesen & Molin, 2003; Luthje, 2005). Franke & Shah (2003) argue that this type of community innovation system works through generalised exchange. In contrast, a market type system would inhibit the exchange of information due to transaction costs associated with attempting to value the assistance ad-hoc,

“it is often not known if a functioning prototype will be developed, if the product will be used by even one individual, if the product will be used by many, and what the value of the product will be for those who use it. In addition, the perception of the individual who has the information and the individual who needs the information might differ.” (2003: 173)

In off-line examples such as mountain bikes (Luthje et al., 2005), canyoning and handi-cycling (Franke & Shah, 2003) the communities come together at sporting event competitions, and users share their modifications with an expectation that others are free to incorporate such innovations into their future iterations. In online

cases such as open source software (Lerner & Tirole, 2002; Lakhani & von Hippel, 2003) the community is enacted on a regular basis through discussion forums and innovations are freely revealed as additions to the “public commons.” This free-revealing with user-communities is also seen in examples of user-entrepreneurship. Shah & Tripsas (2007) documented the process of user-entrepreneurs in the baby-product industry and found that the majority of founders both created and vetted their ideas within the context of a user-community well before the founding of their firms.

2.4 Information flows and innovation

The classic models of product innovation in the firm have considered communication flows both internally and externally (Brown & Eisenhardt, 1995). External parties such as customers and suppliers can be of assistance to the process through controlled “task-oriented” limited communication (Moenaert et al., 2000). Information throughout this process is supposed to be closed off, or optimised through a gatekeeper. A related stream of work in organisational theory on social network analysis has enabled empirical work on structures of information flow (Burt, 1992; Powell et al., 1996). The premise here is that information flows through social ties in a network, which may cut across department or firm boundaries. The combination of disparate information from different contexts can lead to innovation.

The NPD literature stresses controlled processes within the firm, with focused transmission of information between teams or across organisations. In contrast to

the closed firm process models, the user innovation literature is built on the premise of free-revealing innovation information. This enables users to build upon one another's work. In user-communities such as open-source software, lead-users openly exchange information, so long as their designs are non-rivalrous (Baldwin et al., 2006). Intellectual property governance is implemented in the forms of social norms (Fauchart & von Hippel, 2008) or licenses ensuring the community ownership (O'Mahony, 2003). Therefore, openness is important, though knowledge still flows within information regimes.

Throughout the user-innovation literature, the firm is often depicted as an "imperfect agent" conceiving and manufacturing on behalf of consumers and their needs (von Hippel, 2005). Users have local knowledge and this is distributed imperfectly. Innovation comes about through a combination of lead-users willingness to experiment (Morrison et al., 2004) and their incentives to find solutions to local problems. When these users assemble into communities, innovation flourishes through free-revealing and collective assistance (Franke & Shah, 2003).

Firms operating within institutions such as competition and property rights have been engaging user-innovation from different angles. Cast more broadly as "lead-user collaboration", these different approaches range from early stage product concept workshops (Lilien et al., 2002) to creating user-toolkits that enable modifications to be made on the firm's products (von Hippel, 2001; Thomke & von Hippel, 2002; Jeppesen & Frederiksen, 2006).

This debate about openness between firms and user-communities has led to a stream of work under the banner of “open innovation” (Chesbrough et al., 2006). It is similar to both NPD and user-innovation, but differs in that it considers openness from the perspective of the firm. User-innovation is a fruitful literature that has covered examples across industrial products, niche sports, and open source software. It champions the cause of user-hobbyists over the direct market incentives for manufacturers. In NPD, product innovation is seen to renew value-creation for the firm. However, unlike user communities, the firm here is always bounded by efficiency and revenue maximising concerns. Striking a balance between creating something novel and organisational boundaries is an issue for innovation and product development managers. Open innovation has contributed insights on how firms can be more open in working with one another, as well as with user communities.

The user innovation literature pushed us to consider how both the sources of innovation and production can come from outside the firm. This is important in thinking more broadly about the value of users creating products. However, this perhaps over-stresses the production and material aspects of product innovation. Recent work on ‘user innovations in function’ claims (Faulkner & Runde, 2009) that a focus on modification and production by user communities has directed attention away from looking at how communities socially construct innovations. Dahlander et al., (2008: 120) suggest that innovation in online communities is about communication, “through the recombination of ideas, voicing of future product wants and convergence towards a common perception of what is valuable.” Such a statement about ideas points to a perspective of innovation as being about

information, product concepts and consensus on value. This links to the renewed interest within strategy of the co-construction of value (Ramírez, 1999).

Established perceptions of value are the result of a settled niche. The process of settling upon those shared perceptions is necessarily a social one where innovation information is disclosed to a public. The free-revealing of innovation information is an especially important aspect to user-community function. When this information is cast as knowledge, than traditional intellectual property regimes are seen as a barrier. Shah & Tripsas (2007) question how patents may “restrict the volume and quality of information available for all to use as part of the commons.” However, this literature does not ignore governance structures for knowledge flow. Baldwin et al., (2006) concede that free-revealing can only take place in contexts where innovators designs’ are non-rivalrous. O’Mahony (2003) showed how Open Source Software organisations instituted governance structures for knowledge in the “commons.” These works are interesting in how they show innovation arising from information flowing in a somewhat *free* manner. However, there still remains a governance structure, and it is not entirely open or free. This is contrast to the NPD work, which suggests closed boundaries and controlled information flows through a gatekeeper.

There is good reason for a literature using an organisational-level analysis to focus on boundaries and control. If the primary goal is to distill “critical success factors”, then context and contingency theories are inevitably overlooked. Interestingly, the user innovation literature has rejected the boundary and control of information assumptions of the firm and swung drastically in the opposite direction, with the user community employing alternative means of governance. User communities are

presented as the locus of innovation, with firms lagging (von Hippel, 2005). There is a less radical stream within this literature that discusses toolkits and ways for firms to work with lead-users, but the emphasis is still on the user-hobbyists as sources of innovation. Both views consider information flows and sources of innovation, and they both maintain that an organisation is the gravity of activity. This under-specifies the importance of consensus and dynamics of collective action, particularly as it pertains to innovation and the social construction of product markets.

Both the NPD and User-Innovation streams lack empirical structural analyses of social networks. There is an organisational theory literature on innovation that does explicitly consider social networks. In this view, innovation is based on information flow between disparate contexts (Burt, 1992; Powell et al., 1996). Information flows through social ties in a community and access to it becomes a critical question. The network carries private information, and this is novel and timely. Referrals are based on reputation and legitimacy, and can lead to new network ties. Through this mechanism, weak ties can transmit information (Granovetter, 1973). The logic here is that information flow is based on a network and not necessarily an organisational boundary. Innovation comes from diverse forms of information being brought together. This point does not conflict with the logic of NPD and User Innovation, but it is made more explicit here.

Burt (1992) developed “structural holes” to explain how entrepreneurs in product markets who straddle bridge positions between otherwise disconnected networks. Actors positioned this way accrue economic benefits, even if the bridge is a weak tie. This is an important insight for considering information flows and innovation. A dense

network cluster will tend to be homogenous and information will diffuse throughout it. Different clusters will each have their own information saturated throughout. A structural hole will enable this information to transmit across networks. Given that innovation is often about recombining diverse information, the structural-hole is an important in enabling information flow.

Under this perspective, the organisational form of a network can enable timely and quality information flow. Thinking about social networks as conduits for information flow complements both NPD and user-innovation literatures. Organisations structure their activities around information flows (Stinchcombe, 1990). Actors in gatekeeper roles can be structural holes (Burt, 1992) and through their networks, they can bring diverse and limited information into the firm's product development processes. Similarly, user innovation communities can be seen as constituted by network ties. Free-revealing can refer to the extent to which information diffuses in an open manner throughout the network.

The literatures in this section have each put forth a concept of where product innovation occurs. The NPD literature considers the locus of innovation to be within the firm, and the user-innovation literature focuses on the user-communities. Powell et al., (1996) take a different perspective and consider the locus of innovation in the network itself. For an industry like bio-technology, there is a complex knowledge base and expertise is distributed. This industry relies upon inter-organisational collaboration through inter-firm relationships to ensure information flow and learning.

By considering information flow through social ties, the social network view of innovation can be applied to the firm, the user-innovation community and interorganisational networks.

In order to think about information flow through a network, a useful metaphor is a plumbing system. When the pipes are connected, the conduit is closed, and the content can efficiently flow. Podolny (2001) also extended the pipes metaphor to consider how networks may also refract information as a prism. Inherent in this perspective is an assumption that information flows over a platform of closed channels. The network is the platform, and thus information can only flow across the network.

The social network analysis work lets us see empirical evidence of social structures where there is a high probability of information flow. However, this does not account for cases of information leakage or the effect of transmission from low probability sources. Work on innovation clusters is based on a network story, but also refers to there being *mysteries in the air* (Marshall, 1890). Brown & Duguid (2000) explored an ecological view of knowledge dynamics in clusters based on networks and communities of practice. This was based on the classic work from Marshall (1890) that showed the dynamic nature of trade knowledge within the industrial districts of Manchester.

Thinking about an ecological view of knowledge in industrial clusters leads to a loosely coupled model of organisations with sufficient interconnections and movement that the entire industrial cluster can be construed as a type of firm

(Gilson, 1996). This point echoes the Powell et al., (1996) argument about the network being the locus of innovation. However, their model was based on organisations interacting through alliances and collaborative projects. Saxenian (1994) made a cultural argument about Silicon Valley by following Marshall's (1890) suggestion that in localisation, "social forces cooperate with economic ones." She described how knowledge flows intensely across Silicon Valley due to relatively few common cultural platforms such as coffeeshops, schools, and restaurants. Indeed, within a small locality like Silicon Valley, the "level of informal knowledge 'in the air' ... provides an unrivalled key for interpreting the formal knowledge produced there" (Brown & Duguid, 2000: 22). Thus, from a high density of competitors in the same geographic region, there is a substantive amount of informal knowledge flow between employees living there.

Under Saxenian (1994) and Brown & Duguid's (2000) depiction of Silicon Valley, informal knowledge is flowing in ways that are effervescent and not simply due to established social ties. These are not well established information flows transferring proprietary technological knowledge, but they are enabling signals amongst firms in the region. This "almost public" knowledge enables an informal benchmarking that leads to "collective bootstrapping" (Brown & Duguid, 2000). Such an ecological view is only possible when there is a platform enabling potential knowledge flows.

Johnson (2011) described innovation as coming about from serendipity on top of platforms that enable what he calls "liquid networks." This is a dynamic perspective on networks where ties form in a serendipitous manner as a result of nodes being collocated on a common platform.

Work on industrial districts speaks to the dense social networks that form between employees, firms and suppliers, however, as platforms they enables weak ties to form. In thinking about the stability of network structures, social network analysis has focused on cross-sections, and has neglected the dynamics of network formation and dissolution (Watts, 2004). Within industrial districts, there are *mysteries in the air*, where information takes on a dynamic close to free-revealing. The industrial district can be seen as the platform upon which an ecosystem sits. Serendipitous connections are difficult to conceive if flow is private and only occurring along closed channels. Thus, in thinking about *liquid networks* (Johnson, 2011), there is a public component within the platform.

We can think about how the model of private information being transmitted over static networks does not adequately explain a platform where network connections are continually being rewired. The static network model does not adequately explain how weak ties and small worlds (Watts, 2004) enable knowledge to flow between loosely connected social networks. When considering the emergence of a product market, it is analytically tempting to consider it coming about in a static network. Santos & Eisenhardt (2009) illustrate the politics that entrepreneurs engage in nascent markets. They also discuss the cognitive claiming work that is being done by the entrepreneurial firm. Inherent in this story is the issue of connecting to audiences. This shifts the analytical focus to an ecological one. The boundaries of audiences matter less than the enactments of structured webs of meaningful information, or discourses.

Product development has been presented as covering critical processes within the firm, user communities outside the firm, and a social network analysis view that shows how information flows over networks. This activity is focused on the implementation processes of production, but it is also about combining novel information from separate social networks. In thinking about the emergence of a product market, this is social, and clearly an interorganisational activity.

2.5 Interorganisational perspective on product markets

The debate around the movement of information across firms and user communities reveals the broader direction the product development literature has gone in the last two decades. The model of product development focusing on a firm's critical success factors emphasises the perspective of an organisation in relative isolation. Starting from a consideration of information flow being strictly controlled by the firm, one stream has swung drastically in the opposite direction of user communities that freely reveal all. Information for a firm can be a resource, and thus an asset (Barney, 1991). It is then reasonable for firms to maintain trade secrets and the right to disclose innovation information strategically. However, if the firm is instrumentally intent on locking information up and restricting flow, this can hamper product development. Particularly because product markets are social spaces, it is not trivial to prescribe how open a firm should be (Chesbrough et al., 2006). The level of openness may be less consequential when we think about the social construction of a product market. This pertains to the strategy level thinking of discourses in a shared space.

The firm has porous boundaries that enable information flow with external actors. In launching a product in a new space, the focal firm may need to help provide future-looking “tellable and compelling stories” (Simakova and Neyland, 2008) for external constituents to make sense of the new product or technology. However, this is not merely a case of the firm marketing downstream to their external constituents. Suppliers, competitors, regulators, and indeed consumers, in an institutional arrangement can all provide inputs to a product development process. This speaks to the situated nature of product development, in what Van de Ven (1993) calls a “community perspective on the emergence of innovations.” Ultimately, products need to meaningfully tie together concepts and be effectively understood across the community.

Moving from a consideration of product development process by the firm in isolation, we see the competitive dynamic of an industry that consists of a community of producers. The focus on critical success factors can be critiqued as a reductionist view that ignores the fact that producers do not simply make product markets alone. In shifting focus from the firm in isolation to an interorganisational context (Lawrence & Lorsch, 1967), this expressly considers moving from a technology frame to a social one. The firm comes to develop products within a field linking several communities.

One important aspect of moving to a more social level view of new product emergence is that this opens up a perspective for exploring this phenomenon through an interpretivist lens (Porac et al., 2002). Managers have limited cognitive capacities, and they are socialised within firms, markets and industries. Once we consider the product market from the level of an organisational field, the issue

becomes an ecological understanding of resources spaces, public information and niches.

When seen as organisational fields, product markets develop to become local cultures with precedent, orderings and social relations (Fligstein, 2002). Innovation can then be cast as the continual remaking of socially constructed collective frames.

"We tend to think of innovations largely in material terms - jet planes, computers, rockets, detergents, plastic ware - but our concern here is with the establishment of new patterns of concerted action" (Shibutani, 1986: 365)

Then, innovation is about *new patterns of concerted action*. Concerted action is organised through new arrangements of cognitive understandings and forms. These can include new divisions of labour, jurisdictional struggles (Abbott, 1988) and assembling a value network (Brandenburger & Nalebuff, 1995).

Discourse reflects and helps to constitute organisational forms and meaning such as market arrangements. Munir & Phillips (2005) showed how an institutional entrepreneur used deliberate strategies to affect discourse around the meanings of a new technology. This work is important in making the link between the social construction of an innovation and the ability to successfully commercialise a product concept. Discourses are constructed through political struggles and can provide a cognitive scaffolding for innovations to be embedded and meanings-in-use understood by producers, consumers and other market actors.

In thinking about discourse and interorganisational cognition, a collective interpretation of a product can be likened to a cognitive schema (Porac et al., 2002).

Wang (2009) develops an information systems view of information technology being made up of an ecology of innovation concepts. He defines an *innovation concept* as the “collective idea for the development and utilization of an IT innovation.” This is distinguished from the material form that includes physical aspects such as hardware and software, and is explicitly interorganisational,

“While the material aspect of an innovation is always associated with particular organisations, it is the innovation concept that travels across organisational boundaries (Newell et al., 2000).”
(Wang 2009:4)

Wang’s model describes an interorganisational environment where innovation concepts compete for the limited attention of managers. Furthering this, Tsui et al., (2009) proposed that innovation concepts be seen as interrelated in an “idea network.” When considered with an attention-based view of the firm (Ocasio, 1997), these idea networks can be likened to “species in a competitive and symbiotic resource space” (Tsui et al., 2009). This follows an ecological logic of innovation concepts competing as species in a problem domain.

2.5.1 Shared knowledge structures amongst market actors

There has been substantial interest within the innovation literature on shared cognition structures (Orlikowski & Gash, 1994; Dougherty, 1992). This is an idea that combines insights from Organisational Theory (March & Simon, 1958) with those from Cognitive Psychology (Rosch, 1975). Those cognitive units have been given many labels such as “frames”, and “schemas” (Kaplan, 2011) and are often coupled with sense-making (Weick, 1995). Orlikowski & Gash (1994:178) articulated the concept of a technology frame to describe the role of technology in organisations,

“we use the term technological frame to identify that subset of members’ organisational frames that concern the assumptions, expectations, and knowledge they use to understand technology in organisations. This includes not only the nature and role of the technology itself, but the specific conditions, applications, and consequences of that technology in particular contexts”

The technology life cycle literature (Tushman and Anderson, 1986; Utterback & Suarez, 1993) has largely avoided the question of shared frames and cognition, however, this link has been taken on as a research direction (Kaplan & Tripsas, 2008; Porac et al., 2002). Here, in the early stages of a technology lifecycle, frames around its meaning are contested as attempts to define a dominant interpretation. Potential approaches for studying this include how much consensus is required for the interpretation and how long it may remain stable.

Recent work has explored the perspective of shared knowledge structures amongst market actors (Kaplan, 2011; Kaplan & Tripsas, 2008; Porac et al., 2002). The seminal article by Porac et al., (1989) explored how the industry structures in Scottish knitwear were based on to the perceptions of managers of the network of rivals. Porac et al., (2002: 580) considers product markets through a framework of belief systems contained in “collective cognitive structures underlying interorganisational relations.” That is, a network of actors collectively and sufficiently agree upon knowledge structures that interface interaction. This forms a core insight in the cognitive constructionist view of markets. Many of the studies in this literature have focused upon the social structure of the supply side of the market. This has addressed the conception of a market as a field (Fligstein, 1996, 2002) whereby producers closely observe one another (White, 1981) and become legitimated within a cognitive embeddedness (Kennedy, 2008; Porac & Rosa, 1996). A stable market is emphasised through social relations, as being "self reproducing role

structures" (White, 1981). Therefore, cognitive structures are largely meant to indicate social networks of producers in a product category, or industry.

The structured relations of producers helps to constitute a theory of the product market. However, Porac et al., (2002) suggested that every product market has at its core a meaning system that props up the other industry belief systems. They articulated the *product ontology* to describe how communities settled around common definitions of the essential aspects of a product. As cognitive representations, packaging up product attributes, use-cases and buyer characteristics, these knowledge structures are shared by buyers and sellers as a joint understanding that evolves over time.

White (2004) built a model of a market being constructed through discourse and culture in an organisational field. Kennedy (2008) took this forward with an empirical study of a nascent market. By studying co-mentions of producers in press releases and media articles, he found evidence for a "cognitive embeddedness" effect as centrality in the producer network discourse heavily correlated with firm survival rates. As the market matured, this effect diminished. Therefore, in a nascent market, this cognitive embeddedness played an important role in determining the industry definition. As producers were mentioned next to one another, these associations congealed over time into an industry and became seen as meaningful and real.

The maturing of the market occurs as shared understandings assemble, spread and become congealed in social relations. These assemblies and understandings become realised in product artefacts and service processes (Latour, 1987; Pinch &

Bijker, 1987). However, early on, the market is characterised by nascent product ontologies. Porac et al., (2002) analysed how inventors and inventing organisations become critics of extant product ontologies, and in the process start to develop new ones. Market information is defined by intermediaries (Anand & Peterson, 2000) and as the product ontology begins to take shape, the two are intertwined. In extant theory, there is little said about how market information itself takes shape.

2.4.2 Social construction of technology

The social construction of technology literature is helpful for considering the social dynamics around how communities can shape and define a market. The core argument in this perspective is that the source of innovations in technology are impossibly intertwined between developing interpretations and repurposed settled meanings. Instead of latent needs seen as resources worthy of tapping, innovations become defined by interactions and shared meanings held by *relevant social groups* (Pinch & Bijker, 1987). Through a developing rhetoric and discourse from interaction across social groups, problems become defined for the innovation to solve (Latour, 1987). Building from Kuhn's (1970) notion of scientific paradigm, it is proposed that each group develops a shared understanding of needs.

The history of the mainstream bicycle has demonstrated that emergent social groups were formed as a result of the meaning construction process (Pinch & Bijker, 1987). The design we see today is most similar to the design formerly known as the "safety bicycle." Before the stabilisation around this design, there were several designs that

differed with rubber tires versus simply metal rims, the size of the wheels themselves, and the height of the seat. These variations were linked to emergent relevant social groups who were galvanised by concerns such as technical nuisances, traction problems, aesthetics and speed reduction. These groups were not otherwise linked, but for the issues around the bicycle.

2.4.3 The market as an organisational field

The model of a market as an organisational field is an important development in thinking about managerial cognition and looking beyond the firm as a closed system. This speaks to different communities of social actors who come together in patterns of concerted action (Shibutani, 1986). This can be illustrated through thinking about a typical farmers market. These will be held on a particular day of the week at a particular set time in a particular place. Farmers will each have a stall and they will lay out their goods. It is understood by buyers that sellers will announce their prices, and perhaps even lower them just before the end of the day. These are taken for granted assumptions by consumers about the nature of the products being sold and will likely constitute a discourse about *fresh, clean, straight from the farm*. However, these assumptions need to be supported and reinforced by cues and behaviour of sellers.

"A market is a social arena where sellers and buyers meet. But of sellers and buyers to meet, a product has to exist and someone has to produce it. A market depends on the buyers continuing to 'show up' in a particular social space to purchase the product. But the sellers' firms and their status relations define what stability means in the market. They define what the market is about, and their relations define the local culture by which money is to be made and stability produced." (Fligstein, 1996: 31)

The field conception of a market emphasises the network of producers, the local culture understood by both producers and consumers, and a shared understanding of stability. The stability is critical to enable an interfacing platform. Because producers are disparately distributed and initially uncertain about their markets, they cannot securely secure resources (Pfeffer & Salancik, 1978). The set of consumers that make up the demand side of market is also ambiguous and lacking definition. Once it does become defined through a demographic or identity, uncertainty still remains. In order for transactions to be efficient and repeatable, these communities need to coordinate over common standards of commensurability.

Community as a social structure has been central to this view of the organisational field (DiMaggio & Powell, 1983). The reputations of core social actors within the field are granted legitimacy that is greater to those in the periphery. The field is formally defined as,

“those organizations that, in the aggregate, constitute a recognized area of institutional life: key suppliers, resource and product consumers, regulatory agencies, and other organizations that produce similar services or products”
(DiMaggio & Powell, 1983: 148)

The recognition of key actors becomes an essential feature of the field. This arrangement is held in place through the legitimacy of those actors. Another important component in this conception of the field is that organisations produce similar products or services. It is that isomorphism which is important for dominant design theory, although the embedded social structure of community has received more attention within institutional theory.

Intermediaries assist the process through establishing legitimate market information (Anand & Peterson, 2000). This then enables efficient transactions to take place as shared understandings congeal into taken-for-granted industry belief systems.

By standardising interfacing and minimising ambiguity, this enables exchange and trade. Thus, the market as a field is a powerful metaphor that helps to outline social arenas that support common understandings of value. However, this perspective speaks more to stability than it does emergence. There is a gap in this literature explaining the emergence of markets as fields.

2.4.4 Market information

The markets as fields perspective emphasises the importance of shared understandings as stability. Through the social mechanism of legitimacy, market information certifies and provides a platform for these understandings to congeal. As interactions between producers may be limited, this is critical for continued stability of the market. Product categories become the buckets by which market information is organised under standardised definitions that enable commensurability. In this way, market information lets the market know about itself. As market interfaces between producers and consumers are standardised, ambiguity becomes resolved and uncertainty is more controlled.

Market information regimes (Anand & Peterson, 2000) are constructed by intermediaries and are central in the process of market sensemaking. These intermediaries gather, transmit and define information on categories as well as the

rules of the market. The regime “interlocks attention” (Anand & Peterson, 2000) of disparate actors into a simplified cognitive structure that enables relational orderings and rankings. Intermediaries play a major role in constructing a product market.

“using the idea of markets as fields requires one to specify what a market is, who the players are, what it means to be an incumbent and a challenger, and how the social relationships and cultural understandings that come into play create stable fields by solving the main problems of competition and controlling uncertainty” (Fligstein, 1996: 17)

As industry analysts, they are involved in naming and categorisation work (Zuckerman, 1999; Pollock & Williams, 2011). When it comes to new products, they often “provide the first public articulation of” an innovation (Wang & Ramiller, 2009). In this model, intermediaries organise market information into a regime and this enables the field to interact.

Thus, the view of markets as fields can be seen as containing an arena, a network of players, and a coordination mechanism. Under a market information regime, the market information constitutes the field (Anand & Peterson, 2000). Under this conception,

“information typically takes the forms of sales reports, inventory information, trade magazine reports of ‘hot selling’ items, newspaper articles, rumors or gossip with connections to past, present, and future courses of action.” (2000:271).

In the case of Billboard music charts, the perceived legitimacy of a scientific methodology helped stabilise beliefs in product categories as music genres established by the market information regime. This notably changed when a competitor, Soundscan, began offering data that was supposedly more objective. This claim on objectivity was based on a more direct gathering of sales data. As a result of the new form of data, the market shifted its structure. The genre of *country music*, which had previously not been accounted for the by previous method

employed by Billboard, grew quickly as a result of this disruption in sensemaking. Eventually, Billboard bought out Soundscan and incorporated their data-gathering methodology into the market information regime.

What counts as market information is critically important. Rosa et al., (1999) documented the emergence of the Minivan product category by analysing market stories in the form of product reviews magazines and the trade press. They described how sensemaking in the market is a function of market stories,

“For product markets to exist, the cognitive structures underlying the market must be stable across time, extended across space, and shared by many actors on both sides of the market. We suggest that such stability, extension, and sharing is achieved by means of market stories.”
(Rosa et al., 1999: 68)

Market stories summarise experiences actors have with the product. As they are retold and broadcast, large numbers of diverse market actors develop a shared understanding of the product. This is sensible, as stories are critical sensemaking tools (Weick, 1995; White, 1992). However, Rosa et al., (1999) do not outline the political process of how market stories are determined.

Market information regimes are established as a result of claims upon a space (Santos & Eisenhardt, 2009). The gathering, processing and organising of market information is a technology. The standardisation of a market regime is related to a monopolistic hold that an intermediary has upon the claimed space. It is difficult to consider market information within a nascent market. Early market information may not necessarily appear as market stories because the market information regime is not likely to have emerged yet. However, if market information is necessary for the

product market to stabilise, then we have a conflict of orderings. Therefore, we need to consider other means of how market information may be assembled and used.

2.4.5 Early market information as rumour

Product markets as fields emerge once shared understandings have stabilised (Rosa et al., 1999) and uncertainty has been controlled (Fligstein, 1996). Meyer et al., (2005: 467) suggest “field configuring events” are critical gatherings that effectively constitute a starting point, as “microcosms of nascent technologies, industries, and markets.” This definition is based on the premise that actors from diverse settings will come together in an event with the purpose of constructing a field. Coming together in such an event constructs networks, builds reputations, and may set standards going forward. This conception is aligned with the view of field stabilisation coming from controlling uncertainty. However, this speaks less to resolving ambiguity in the early moments of fields. How would a field configuring event be organised in the first place? How would the actors and issues be assembled into an event where main purpose is to define a field? Considering these questions is more about ambiguity than about uncertainty.

Organisational theory describes ambiguity as the lack of clarity about meaning and the implications of particular events or situations (March, 1978; Weick, 1995). In a nascent market, this is an unstructured setting with extreme ambiguity that stems from a “lack of recurrent, institutionalized patterns of relations and actions” (Santos & Eisenhardt, 2009). This differs from uncertainty, which is about lack of predictive

power. Legitimate market information helps to resolve ambiguity as the field settles around cognitive units as product categories. Early market information can begin to come into view as ambiguity is being resolved into a market information regime.

In considering where early market information might be found, a starting point is that fields may assemble around varying levels of objectivity. The framing of market information is likely to be objective and scientific, but this is not always the case.

Anand & Peterson (2000) suggested that market information can be found in several sources.

“The methodology used for framing market information is vital to the social construction of a market. Thus, for example, while some fields may develop traditions of market information assembled according to scientific methods of sampling and analysis, other fields make do with information that is little better than gossip and rumor” (Anand & Peterson, 2000: 272)

Thus, market information in some contexts can come from rumour or gossip about future courses of action. In the context of a nascent market, rumour is particularly interesting to consider further. Without a community structure, the market as a field is inchoate and ill-defined. In the context of extreme ambiguity, this is a situation conducive to rumour formation (Shibutani, 1966; Rosnow & Fine, 1972). From a pragmatist perspective, it may also be the only information about the future market available.

In its basic form, a rumour is a proposition, developed without secure standards of evidence (Allport & Postman, 1947) or a reliable sponsor (Fine & DiFonzo, 2011). It represents discussion about future states that people care about and the content is situated along shared schemas that are held by a social context. Rumours tend to emerge in situations with ambiguity, uncertainty and interest. Allport & Postman

(1947) developed a law of rumour, $r = a * i$, where the strength of a rumour is equal to the level of ambiguity multiplied by the level of interest. The sociologist Tomotsu Shibutani (1966) later developed a related theory of rumour as collective sensemaking entitled “improvised news.” He showed how rumour was a collective transaction in which people “caught together in an ambiguous situation attempt to construct a meaningful interpretation of it by pooling their intellectual resources” (Shibutani, 1966:17).

The component parts of rumour are cognitive and they serve to communicate activity. The content is situated in local schemas and derived from shared beliefs (Prasad, 1935). It is distinguished from other forms of communication like gossip, which is localised as idle talk by groups and usually about people and their reputations (Shibutani, 1966; Rosnow & Fine, 1976; Burt, 2008). It shares with gossip a low degree of formalisation. However, rumour is more concerned with meaningful propositions about a future state that concerns a public. In this way, rumours are hypotheses about how the world works that allows people to handle change amidst the uncertainties of life (Rosnow and Fine, 1976:30).

As an “information black market” (Kapferer, 1990), there is recognition of the sense-making function of rumour and that rumours can themselves be substantially true as well as false (Bordia et al., 2006). Recent work from Bordia and DiFonzo (1998, 1999, 2004) has explored the sense-making aspect of rumour on the Internet. Their research supports the notion that rumour is a form of social cognition; groups engaging in a sense-making process through discourse. One such study (2004) focused on how a series of computer-mediated forums used rumour as a form of

group problem-solving. The outcome of the study noticeably matched the collective sense-making conception (Shibutani, 1966).

Although rumour has been described as often constituting false information (Sunstein, 2009), it can at times also become confirmed as true. Fine & DiFonzo (2011) more appropriately describe rumour as “uncertain knowledge” that lacks an institutional sponsor. This corresponds to Shibutani’s (1966) model of rumour being “improvised news”, as a collective transaction that emerges in situations of ambiguity and uncertainty. In this pragmatist tradition of sociology, ambiguity has been described as the result of a breakdown in institutional channels of information. We can invert this and consider rumour as occurring in the emergence of such institutional channels in a field.

This model of rumour may be extended to the early moments of nascent markets as social spaces use developing early market information. Nascent markets are situations of extreme ambiguity, as they lack institutional definition or precedent (Santos & Eisenhardt, 2009; Aldrich & Fiol, 1994). Market information regimes (Anand & Peterson, 2000) may be the attracting force in the early moments of a market. Institutions, social relations and structural forms will come to define the market under a logic and legitimacy embedded in reputations, but early on, information in the form of rumour may constitute an early field.

2.4.6 How product markets emerge as nascent markets

Despite the importance of building new markets to innovation and entrepreneurship, research in this area has only received relatively minor attention. The technology life-cycle is an evolutionary economics theory that has proven rigorous across many contexts, but its scope is quite limited to economic variables. The sociology of markets has matured as a literature but is still fragmented in sociological explanations of new product markets (Fligstein & Dauter, 2007). Both perspectives have scantily focused on the early moments of markets, but there has been some work in this area. One frame for considering market emergence has been in nascent markets (Santos & Eisenhardt, 2009; Aldrich & Fiol, 1994). The deficiency in researching this phenomenon may be due to the difficulty in gathering data, as it is often ephemeral and rife with ambiguity over where to look (Forbes & Kirsch, 2011). It is likely that this also may be due to an epistemological bias in seeking out stable phenomena. As a result, research on the emergence of product markets has been understudied.

Nascent markets are those settings in early stage formation, lacking definition amidst extreme ambiguity (Santos & Eisenhardt, 2009). With unclear product definitions (Hargadon & Douglas, 2001) and little industry structure (Eisenhardt, 1989), it becomes difficult to pin down what is what, who is relevant, or where the boundaries are located. An associated line of reasoning is that of nascent markets resembling emerging fields (Fligstein, 2002; DiMaggio & Powell, 1983). This thinking, borrowed from institutional theory, holds that legitimacy is a primary mechanism sustaining social structures. In a nascent market, reputation-based or institutionalised

legitimacy will not yet be present. This corresponds to the call from Forbes & Kirsch (2011) to investigate the phenomenon using macro-level literatures that cover economic, ecological, institutional and cognitive perspectives.

The cognitive construction view of markets is underpinned by the proposition that product markets are networks of producers and buyers linked by shared assumptions and frames of reference (White, 1981; Porac et al., 2002). It describes socially constructed market categories as the outcome of producer and buyer sensemaking and how this interface provides a stable basis for repeated transactions (Porac et al., 2011; Kaplan, 2011).

Santos & Eisenhardt (2009) created a framework for understanding nascent market construction that delineated *claiming* as cognitive boundary construction activities, *demarcation* as micro-political activities and *control* as more traditional competitive strategy activities (Porter, 1980). The claiming activities are cognitively derived and most closely link to work on product definition, collective technology frames (Kaplan & Tripsas, 2008) and product ontologies (Porac et al., 2002). This is helpful in considering a potential sequence of how a product market emerges from a nascent stage.

2.4.7 When product ontologies emerge

The existing work on product ontologies suggests that they come about from the efforts of an inventor or inventing organisation acting to solve design problems (Porac et al., 2002). Later, inventors actively push them out and they evolve through the stories across buyers, sellers and other key market actors. This is reflected in market discourse as a public conversation, with stories, texts and other documents. For firms innovating in a nascent space, they must actively contest boundaries (Santos & Eisenhardt, 2009) and firms and institutions associated with existing markets. For a producer, this means resolving the ambiguities around figuring out which other producers are in the same space and what to produce. For customers, the emergence of new product markets challenges their understanding of products, matching cases to contexts, and assigning value (Rindova & Petkova, 2007; Rosa et al., 1999).

In the early stages of market emergence, there is a lack of definition, precedent, and a dominant logic to guide strategic actions (Aldrich & Fiol, 1994). In inchoate, or nascent markets, there is extreme ambiguity (Santos & Eisenhardt, 2009).

Legitimacy is an important mechanism used to construct the scaffolding necessary to claim and demarcate cognitive referents. When market 'mediators' such as the media mention firms together in discourse, this legitimacy helps to signal in the emerging organisational field of which make up the producer network (White, 2004).

Market interfaces are socially constructed as an outcome of joint sensemaking by producer and buyer communities (Rosa et al., 1999; Abrahamson & Fombrun, 1994;

Porac et al., 1989; Kaplan, 2011). Which producers are to be included in the definition of the industry network (Porac et al., 1995; White, 1992) is coupled with industry meaning systems that also define the product. This becomes taken for granted and embedded in a belief hierarchy over time (Porac et al., 2002). Santos & Eisenhardt (2009) suggested that the initial activity of entrepreneurs was to claim the cognitive space and become a referent. This view speaks to the identity of the firm, perhaps as a brand. The product ontology focuses on the micro-level of a space seen through the essential aspects of a product definition.

One claim here is that the early moments of a product market will be characterised by nascent product ontologies. Porac et al., (2002) discussed in their review of this work that product ontologies emerge from inventors and inventing organisations who become “ontological critics.” Once they develop designs based on these nascent product ontologies, producers release products and market stories begin to be told as consumers use them. Thus, new product ontologies are initiated from the efforts of an inventor or inventing organisation acting to solve design problems. Later, inventors actively push them out and they evolve through the stories across buyers, sellers and other key market actors (Rosa et al., 1999). This is reflected in market discourse as a public conversation, with stories, texts and other documents. Market discourses are also affected both by intermediaries and other actors in the field. Legitimate market information is defined by intermediaries as the product ontology takes shape.

Considering these issues, the question of product market emergence is a multifaceted one. First, there is the question of genesis; in referring to the product

market, when can we speak of a *nascent market*? How early in the emergence is it possible to conduct a tractable analysis? Within these questions is the issue of how to sensibly analyse what is stabilising. The field conception of markets speaks to social "structures [that] involve both cognitive understandings and concrete social relations" (Fligstein, 2002: 32). Although two elements are mentioned, this work has focused attention on social relations. It says that producers as institutional actors become socially embedded (Granovetter, 1985) in sets of rivalry (Porac & Thomas, 1990), whose definitions are facilitated by intermediaries. Cognitive understandings are thus a residual grouping worthy of peering into.

Second, there is the question of the space where the product market gets created. Product categories are shorthands for claimed spaces containing cognitive referents (Santos & Eisenhardt, 2009), industry reputations, and performance metrics (Christensen, 1997). The politics of establishing the rules of the game are embedded in settled and reproducible role structures (Fligstein, 1996; White, 1981). The implication of this view is that the individual firm is embedded in a social environment of a field and is subject to norms, culture and path dependence of other institutional actors. The product category is a conceptual system (Rosa et al., 1999) that is underpinned by a product ontology (Porac et al., 2002). The sensemaking process of establishing a product ontology occurs amidst market stories. Therefore, in the absence of a stable set of social relations, an analytical strategy for observing product market emergence is to find a public that is deliberating the pre-history of a product ontology and to trace the discourse as it forms into a meaning structure.

2.5 Towards a discursive view of product market emergence

In this section, I begin to develop a preliminary model describing early moments of nascent markets by suggesting that a public can deliberate discourses and begin to settle around associative forms. This echoes Hoffman's (1999) model of organisational fields as a constituency of key actors based around issues and discourse, not markets or a static community structure. The analytical difference here is that a nascent market may contain both big established firms and small startups. In a similar view of emerging organisational fields affecting entrepreneurial agency, Garud et al., (2007) argue that shared discursive symbols are used as "resources to formulate and legitimize heterogenous visions". They consider ante-narratives (Boje, 2001) as incomplete narratives that help to recontextualise forms. As part of the sensemaking process, this enables actors to draw upon cultural narratives to reshape fields. I propose that ante-narratives are similar to rumours and speculation. Extant theory suggests product ontologies begin as market stories that are told after producers release products and buyers attempt to make sense of them. Moving the lens to incorporate rumours and speculation shifts the perspective to temporary arenas of collective behaviour where social ties may be fleeting and cognitive forms come about from a slightly different type of sensemaking. The main theoretical claim here is in line with the assertion from relational sociology that discourse and cultural forms are a dually ordered (Kirchner & Mohr, 2010). The product market is indeed a field made up of actors tied together through a web of social relations. However, it is also underpinned by a product ontology as discourse. This can take on the form of a meaning structure through a common nomenclature. It is the structure of this discourse which I theorise on.

2.5.1 Organizing vision

Thinking about fields as being constituted through discourse has been developed in a related literature of Information Systems. The work on the IT “Organizing Vision” (Swanson & Ramiller, 2007) refers to the process of communities of institutional actors who collectively define a discourse around an emerging technology. These institutional actors are supported by an industry that write, and sell ideas about technology (Wang, 2009). The discussion of a technology innovation occurs early on as the *organising vision* is developed. This activity is conducted by consultants, and industry experts as,

“the organisation is not alone in its struggle to make sense of the technology. Instead, it belongs to a complex community of organisations, the many members of which actively contemplate new technology and ponder, to varying degrees publicly, what it means and where it is going” (Swanson & Ramiller, 2007: 459)

This *organizing vision* provides a theoretical base for analysing the discourse and diffusion that shapes both the development of technology and its use in practice. It embraces features from the innovation diffusion literature (Rogers, 1995), sense-making (Weick, 1995) and institutional factors. Not only are institutional processes at play from the beginning of the diffusion, but also as the innovation is being developed. Through conferences, trade-shows, and the emergence of buzzwords, the interacting communities of vendors, consultants, trade associations, and customers develop a rhetoric and vision about a new technology that is a discourse. Within an interorganisational community, social actors position their multiple interpretations and seek to legitimate them within the collective vision. This speaks to the shaping of meaning of a technology as a social discursive process.

2.5.2 Product market as discourse and product ontology

Porac et al., (2002) developed the *product ontology* to explain a shared cognitive understanding in a product market. The product ontology is an understanding of a product category that links the artefact, its understood attributes and usage conditions. A product market is a cognitive space that contains related but partitioned product categories. Such a space is a multilevel knowledge structure that brings together a constellation of attributes with a social network of producers (White, 1981), organisational reputations, and industry recipes. This framework helps to explain market actors and contexts by focusing on "collective beliefs that influence the interpretations and actions of organisations competing within particular industry segments" (Porac et al., 2002). These collective beliefs subsume ideas of technology in practice (Orlikowski & Gash, 1994) and a market as a social arena (Fligstein, 1996). Most importantly, the product ontology is described as a market-core that sits as a base upon which industry beliefs stand. A coherent product ontology also enables commensurability, which enables producers to be judged comparable and members of the same space.

Rosa et al., (1999) propose that product markets are underpinned by conceptual systems that are developed in market stories. In their study on the Minivan, they showed how it became commonly understood through the constellation of attributes that include "car-like handling", "front-wheel drive", "cupholders", "low step in height", "passenger side air bag", and "storage large enough for a 4 x 8 piece of plywood" (Rosa et al., 1999). In this work, the array of known producers, product models and other essential features of the product market were acknowledged, but

not explicitly theorised. They traced the evolution of a product market category that emerged from a dialogue about product similarities and differences that took place in the public space of mainstream media.

Market stories were seen as part of a sensemaking process that helped social actors “build the base of tacit knowledge that supports the market in question” (Rosa et al., 1999:70). The volume of market stories was presented as evidence of schemas settling around a constellation of attributes that made up the minivan category. These conceptual belief systems then serve as a market core that acts as a platform for other industry belief systems such as boundary beliefs, industry recipes, and reputations. Together, these belief systems constitute an *ontology*.

“Since the manner in which people coordinate their efforts or oppose one another rests on their definition of objects, any alteration of custom or law involves some change in outlook. What is actually displaced are the expectations with which people address themselves and one another as well as what they take for granted as natural. Thus, social change is basically a mental process; it involves the transformation of the perspectives of the people involved.”
(Shibutani, 1986: 365)

An ontology refers to what counts as reality, or is taken-for-granted as being natural. Social ontologies are “systems of categories, meanings and identities within which actors and actions are situated” (Ruef, 1999). That is, a common inter-subjective recognition within a social context of categories that each are imbued with meaning and identity. There is a history of research into ontologies in Anthropology and Sociology, as societies and institutions generate category systems organising the material world. Categories get constructed as essentialist definitions, and make up category systems. These are then used to fit material things meaningfully into an organised understanding that is mutually exclusive and collectively exhaustive (Douglas, 1987; Durkheim, 1915). In the neo-pragmatist turn of social science,

language comes to form labels and discourses which enact a legitimate recognised reality for a social context (Berger & Luckman, 1967).

When applied in the product market context, an ontology is the cognitive structure that encapsulates the cues and essential conceptual elements of a product category. Interestingly for social science, ontologies emerge in context and become firmly held in cultural milieus as *meanings-in-use* (Porac et al., 2002). Empirically, social ontologies become common sense for those actors within a social-context. It is akin to what Zerubavel (1999) calls a “thought community.” Porac et al., (1989) studied the Scottish Knitwear industry by asking managers across firms who they felt their competitors were. The results found clusters of producers whose products also corresponded to sets containing shared constellations of attributes. The market was partitioned and space was understood both by which producers were involved and by an array of attributes contained in the artefact being sold.

A cognitive structure coming about from discourse speaks to the stabilisation of social meaning. This reflects a linguistic turn in social science where cultural understandings can be systematically measured. Mohr (1998) has lead a nascent body of work in relational sociology based on measuring meaning structures in discourse using quantitative methods. A core insight from this work is that social structure and meaning structures are mutually constituted (Kirchner & Mohr, 2010). This allows us to think about discourse beyond static social networks and diffusion information flow, and rather into measuring meaning structures according to how discourses are enacted.

Social network analysis offers a cultural perspective on network clusters tending towards homophily. Within organisations, different departments settle upon collective cognitive frames as “thought worlds” (Dougherty, 1992). Successful innovation comes about when the boundaries can be overcome and diverse views are combined in novel combinations (Carlile, 2002). An assumption in this view is that a boundary encompasses a network cluster that is homogenous in its cognition and frames. In this model, when actors are sufficiently connected, they converge towards a common frame.

The model of communication based on social networks is sensible in many cases, particularly with the organisational structure of community. However, other models of communication may not be as limited to networks that enable information flow. Broadcast models with mass media and small-world networks are examples of communication that do not necessarily require community to diffuse. The model of informal knowledge dynamics in Silicon Valley (Saxenian, 1994; Brown & Duguid, 2000) described earlier in this chapter speaks to the “almost public” nature of information. It is not fully clear here how information flows. In some cases, it may be accidental leakage, in others, it may come from shared informal social spaces. However, in order for this informal knowledge to flow through ambiguous channels at a critical mass, this is contingent upon a platform such as the locality of cultural spaces. Johnson (2011) develops a view on innovation platforms based on Jacobs (1992) urban model of sidewalks enabling serendipitous interactions. It is possible to bumping into somebody on the sidewalk without planning to do so. The same cannot be said for driving in cars. Therefore, certain cultural spaces can enable “almost public” information flow.

When we shift from a static structure of social networks enabling information flow, this can move thinking into spaces that can serve as platforms (Johnson, 2011; Adner, 2012). Studying such activity on a platform is less about seeking a strict boundary, and is more about tracing the presence of types of activity. Instead of looking at social structure as the conduits for information, it is possible to study a platform and find structures of meaning in the discourse itself. The work of Mohr (1998) lends itself to studying cognitive structures as forms of social meaning. This suggests that social structure and meaning structures are mutually constituted (Kirchner & Mohr, 2010), however, it may even be possible that one slightly precedes the other. Thus in order to find the emergence of a social structure, it may be possible to find a meaning structure emerging in discourse first. To find such discourses there would need to be a platform enabling it. Either this will occur in existing social structures, or more temporary spaces such as publics.

2.5.3 A sociological perspective on publics

As outlined in the sociology sub-discipline of collective behaviour (Goode, 1992), publics are discursive arenas, similar to discursive-fields (Hoffman, 1999; Snow, 2004). Dewey (1927) described how a public is made possible through mass communication. This is related to the *public sphere* (Habermas, 1992), but is not limited to geographic locality. Shibutani (1966) reckons that a public is seldom organised, however it does have an evolving structure, contingent upon the organisation of communication channels. Mische (2009) defines publics as

"interstitial spaces in which actors temporarily suspend some aspects of their identities and involvements in order to generate the possibility of provisionally equalized and synchronized relationships. These spaces buffer relations between individuals and collectivities that otherwise may be engaged in particularistic and contending projects. They often draw upon ambiguity and ritual to find points of connection that generate productive relationships and new forms of joint action". (2009: 21)

This definition emphasises the discursive interactional notion of the public as a space. The temporary gathering of social actors in a discourse *draws upon ambiguity* to generate *new forms of joint action*. The idea of a public has long been established and yet is instrumentally absent from many literatures such as organisation theory and innovation studies. In the tradition of pragmatism, Dewey (1927), Blumer (1948, 1966) and Shibutani (1966; 1996) dealt extensively with the concept, elucidating social behaviour that it is neither an organisation, a community or a crowd. In one conception, classic sociologist Gabriel Tarde contrasted two different social "collectivities",

"one of a public and the other a crowd, the public is a modern creation and exists only by virtue of the mass media; the crowd has existed wherever and whenever humans have congregated in substantial numbers. And while a given individual can – and always does – belong to several publics simultaneously, one can belong to only one crowd at a time" (Tarde 1898/1969, cited in Goode, 1992: 21).

Tarde's observations were historically situated in a time when French aristocrats looked out their windows to find riots in Parisian streets by seemingly unruly and chaotic crowds. Just after the French Revolution, society had been dramatically disrupted and was being reshaped. Tarde and his contemporaries such as Emile Durkheim worked to compile systematic and scientific insights in early attempts to develop the discipline of sociology. For the distinction between publics and crowds, the former was seen as a rational, deliberating group connected by, and being aware of, social and political issues. Publics were later positioned as important to a

functioning democracy, as well as preceding the development of a state (Dewey, 1927)

A public is an arena for the development of propositions as uncertain knowledge. Rumours flow through a public attempting to make sense of a situation that is interesting to them, ambiguous and uncertain (Shibutani, 1966). Considering the market information regime (Anand & Peterson, 2000) argument about information constituting an early organisational field, I further concentrate a focus on defining *product publics*. As loosely tied collectivities, these interaction contexts deliberate and develop discourses around products. These temporarily constituted collectivities are not communities, but are made possible through mass-communication, particularly social media. The outcome of activity in a product public can be seen as “collective technology frames” (Kaplan & Tripsas, 2008), or a product ontology (Porac et al., 2002).

I propose here that the *product public* is primarily a discursive group that gathers in the pre-history of a product market. Tarde described a public is a kind of ‘spiritual collectivity, a dispersal of individuals who are physically separated and whose cohesion is entirely mental’ (Tarde 1898: 277). The product-public proposed here is not constructed through strong social ties, but is a temporarily constituted interaction context where discourse is deliberated as rumour and speculation. Legitimacy of social actors is not the main mechanism of operation. In line with Shibutani's theory of rumour as collective sensemaking, this is a pragmatist view where the product-public is an interactional space that develops 'technology frames' and expectations. The public is only made possible through mass media. Before a product market

emerges as a field, a temporary grouping of people become sufficiently interested in developing a discourse. Such activity will entail temporarily suspending some aspects of identities to engage in a rhetorical exercise in a space enabled by social media. Therefore, a product public can be seen to come together in temporary, loosely bound settings to develop a discourse.

2.5.4 Studying the emergence of a product ontology in a product-public

The product ontology work was an attempt to describe the base of a framework for the layers of market belief systems. This helps us to think about what underpins product markets. The cognitive structure includes associations such as an array of attributes and use-cases. Thinking in this way also brings up the question of the early moments of a nascent market. *How do cognitive associations form and where do they develop?* The markets as fields view potentially over-stresses the logic of social ties and community in explaining market emergence. The nascent market work describes how social networks are not present initially and form over time. I propose that product ontologies develop early on in product publics.

An ecological logic is important to consider in the shared discourse of an interorganisational environment (Rao, 2002). Although labels can indicate organisational fields settling (Grodal, 2007), during an era of ferment, they can be in a state of flux and quickly replaced. In this call for a discursive view of product market emergence, I attempt to operationalise the product ontology by considering the meaning structure that is formed.

By moving away from a purely material consideration of product markets, we are able to think about the product ontology as a conceptual system of meaning. The key questions in this is are of emergence and stability. Rosa et al., (1999:71) suggest that an important outcome of developing a conceptual system is an “agreement on a set of core elements that define the concept.” They develop a sensemaking process whereby market stories are the vehicles by which this occurs. Thus the volume of market stories are seen as an indicator of a product ontology’s stability, and the content shifts over time to focus on specific models and attributes. This is particularly interesting theoretically, as Rosa et al., (1999: 70) claim that market stories have predictive power in “anticipating what the core attributes in emerging product markets will be.” Market stories constitute the discourse that makes up a social construction of a product market, but in the early moments, the discourse may be a meaning structure of cognitive associations.

Emergence is about when complexity collapses to the point where a level of simplicity results in the recognition of something.

“Emergence refers to the ways in which chaos and infinite complexity may evolve new levels of order and relative simplicity” (van Ginneken, 2003: 117)

As opposed to the mechanistic logic of determinism, emergence follows an ecological logic of probabilism. In the early days of a product market’s emergence, it is a nascent social space, rife with extreme ambiguity (Aldrich & Fiol 1994; Santos & Eisenhardt, 2009). Once stabilized, a product market is said to settle into a hierarchy of shared belief systems (Porac et al., 2002), that includes a known network of producers (White, 1981), and definitions of value and dimensions of commensurability. This boundary work is a social achievement, in so far as the

interorganisational cognition becomes distributed and eventually blackboxed (Latour, 1987). Relatively decoupled social actors through a series of social mechanisms come to share understandings of market definitions and reputations. This understanding is often taken as common-sense and orderings can affect value (Benjamin & Podolny, 1999; Ramirez, 1999).

2.6 Finding product publics online

Nascent markets are said to be inchoate and are difficult to study (Geroski, 2003). Data is often ephemeral and by definition they are ambiguous and lacking articulation (Forbes & Kirsch, 2011; Geroski, 1993). They are seemingly created alongside emerging organisational fields (Aldrich & Fiol, 1994; Grodal, 2007). Earlier in this chapter, I explored how nascent markets are organised in part by information intermediaries constituting fields (Anand & Peterson, 2000). In social media, collectivities of technology blogs regularly explore product propositions well before companies make official announcements. The field configuring event (Meyer et al., 2005) is one sensible claim for the genesis of a nascent market, however, it may not always be clear what that event will be. There also may be cases where a field configuring event is preceded by a pre-history where a discourse is developed. This matches the condition for studying product publics engaging in discursive activity before the product market has stabilised.

Online communities have become the ubiquitous framework for scholarship on organisation in the online space. Despite a rich portfolio of options from those who

study groups and collectives, *community* has been settled upon as the central theme (Preece & Maloney-Krichmar, 2005). In innovation studies, there has been a recent surge of interest in online research with community as a key concept (West & Lakhani, 2008; Dahlander & Frederikson, 2011). With the Pareto skew of online participation being a robust empirical result, many of these studies construct networks as core-periphery structures. In such a structure, the core group of users is seen to possess strong ties, where reputation and trust hold the community together, with the periphery being more ephemeral at the boundary. Little is said about whether there are substructures within the periphery and how involved the core is in discursive activities. In order to find product publics online, it becomes necessary to think about contexts where activity is not regular and stable like a community, but is topically organised.

Some work in online communities has looked at discussions as networks (Gonzalez-Bailon et al., 2010). This considers information flow in communities occurring through discussion threads similar to how emails are linked by senders and recipients (Kaltenbrunner et al., 2009). Such a view emphasises that the structure of the network shapes how actors are exposed to information.

Faraj et al.'s (2011) called for Organisational Theory to consider ways of studying knowledge and flow in online communities that embrace fluidity of membership and participation. This can be usefully combined with work from relational sociology (Mohr, 1998) on meaning structure to study both discourse and interaction space. A fluid online community is seen as one where boundaries, norms, participants,

artefacts, interactions, and foci continually change (Faraj et al., 2011: 3). A definition used to motivate this concept is,

“open collectives of dispersed individuals with members who are not necessarily known or identifiable and who share common interests, and these communities attend to both their individual and their collective welfare” (Sproull and Arriaga 2007)

Such fluidity with the flowing in and out of resources behooves researchers to study dynamics of ideas rather than simply structure. This also opens the door to a consideration of online communities under the reconceptualisation of publics.

A public is tied by common interests, does not have a static structure and members may not know each other (Dewey, 1927). Several interesting features arise when studying how online publics collaborate on knowledge generation. When there is anonymity and social disembodiment, or decontextualization of ideas, they become independent of their authors and original context in which they were written. It then becomes possible for the publics to deliberate in a process that,

“allow[s] combination and recombination of ideas unconstrained by conventions of social interaction, shared context, or even a shared set of ontological assumptions” (Faraj et al., 2011: 6).

Thus, by temporarily suspending some aspects of identities, members of publics may engage in a form of deliberation that does not require continuous interaction. The outcome of such a deliberation can be a stable meaning structure that constitutes a nascent product ontology. Whether this forms a stable meaning structure is an empirical question that requires further investigation.

In the fluid approach to online collectivities research, the actions and intentions of users are resources that affect the conversational dynamics (Faraj et al., 2011).

They can include time, passion, socially ambiguous identities, social disembodiment of ideas, and temporary convergence. As the resources can flow in and out, this causes dynamic tensions with positive and negative consequences for knowledge collaboration. The way these tensions can be overcome by the collectivity defines its generative ability to develop knowledge.

On tech blogs, there is a useful distinction to be made between the “front stage” and the “backstage”. Many of these sites are written by editors who publish content on a homepage that is similar to a magazine. There are also comment sections and sometimes forums where users are able to contribute commentary. The front pages are most accessible and easiest to find, and thus constitute a *front stage*. The latter is more diffuse, harder to find and constitutes a *backstage*. These divisions map onto the Goffman (1959) view of social life as performance. Front pages of tech blogs are clean, accessible, and standardised. Comment sections and web-forums incur transaction costs in finding and organising content. As such, they are less accessible and less standardised. They constitute a *backstage* that is messy, complicated and not easily accessible to newcomers or outsiders.

The logic of a *front stage* corresponds with the components of a product public. With relatively minimal effort, a tech blog can engage in deliberation around the speculation of a product market. This can get picked up by a web crawler aggregator such as Google, Slashdot, or TechMeme, and the message amplified (Droge et al., 2010). Conversely, a *backstage* requires an investment of time and effort to track and follow content that is much larger in scale and less organised. The logic of the

backstage corresponds to a community where engagement in such activity requires a level of commitment.

In looking for product publics empirically online, the sensible approach would be to consider the dynamics of discursive activity to find traces of organising principles. Publics and communities are contrasting organisational forms. Without survey or interview data, it is difficult to infer social network ties from online discussion data. Faraj et al. (2011) asserted that online communities have been studied with the static frame looking at issues of governance, user motivations and their relative positions within a structural form. They call for research into “flow and connections of ideas over time” (Faraj et al., 2011: 12). Although methodologically challenging, this holds promise for analysing discourse and knowledge collaboration. This is in contrast to work in social network analysis that has focused on cross-sectional slices of network structures, often by tracking centrality and structural equivalents of actors’ positions (Kilduff et al., 2006; Borgatti et al., 2009). However, both approaches could be useful in analysing online discussion data. In order to distinguish between public-like activity and community-like activity, it would be necessary to study discussion dynamics of timing, content, and user activity. Users who come together for a short period of time to discuss a topic would demonstrate public-like dynamics. When users show commitment to a particular space and engage in discussions across topics, this would demonstrate community-like activity.

2.7 Studying the emergence of a product ontology

Along with the question of which actors are embedded in the producer network of a nascent market, it is important to consider what forms become embedded in its cognitive structure. Rosa et al. (1999) documented the emergence of the minivan product category through a public conversation between producers and consumers. They showed a cognitive “knowledge structure” settling, linking the artefact, its understood attributes and usage conditions. Porac et al. (2002) bundled this in the “product ontology”, and reframed it as cognitive representations packaging up product attributes, use-cases, and both buyer and seller characteristics. These knowledge structures are shared by buyers and sellers through an iterative joint sensemaking process over time. The product ontology is the knowledge structure that forms the category (Porac et al., 2002). It is a package that constitutes the market schemas, with product categories as an outcome of interorganisational sensemaking. Mohr (1998) outlines that social structures and meaning structures are mutually constituted in discourse. A nascent product ontology can potentially start to develop in a product public. This may help to explain the early moments of product markets before a producer network stabilises.

In considering the product public coming together online, the product ontology is developed as a form of uncertain knowledge (Fine & DiFonzo, 2011). As it forms into a coherent discourse, it will take on a structure of meaning that can be systematically analysed (Mohr, 1998). The product ontology is a market core that is dynamic. Extant theory says it evolves as producers come into the market and consumers use the product. It is embedded within early market information as social actors attempt

to make sense of the extreme ambiguity in the space. Market stories such as product reviews in magazines have not yet appeared because there is nothing to review.

Product markets have been developed in this chapter as social constructions that serve to cognise interactions of groups of social actors. Their emergence is the result of political contests among producers (Fligstein, 1996), and intermediary market regimes organising a field through information brokering (Anand & Peterson, 2000). Within the broad product market, industrial structures come about from firms finding differentiating niches (White, 1981). The growth of these niches are facilitated by “cognate market information” (Anand & Peterson, 2000: 282) and settle into product categories. Each niche is contained under the broader logic of the industry, however, it does develop more localised logics such as performance or dominant design. There is a conceptual link in this work to the stabilisation of dominant designs (Kaplan & Tripsas, 2008), but it has not been fully developed.

I proposed here in this chapter that nascent product ontologies begin to develop earlier than this in discourse as a form of incomplete market stories. As Garud et al. (2007) developed in a discursive view of emerging fields, ante-narratives (Boje, 2001) as incomplete narratives that help to recontextualise product markets. Ante-narratives are texts with minimal narrative properties. They can include slogans, technical or product descriptions, or brief press releases, and have some rhetorical power through links to broader discourse. Garud et al., (2007: 9) build a view of entrepreneurs using ante-narratives to “draw on to frame and make sense of their activities.” As a product ontology develops, it is interesting to consider how

participants in a product public may employ ante-narratives to help frame a product ontology.

An analog to work on product ontologies comes from work on emerging organisational fields. Grodal (2007) studied the emerging organisational field of nanotechnology by looking at labels as the outcome of meaning. Santos & Eisenhardt (2009) studied several nascent markets, following the strategies of entrepreneurs. Aldrich & Fiol (1994) laid out several proposals based on ideas of cognitive legitimacy of social actors. There is a gap in work that directly studies product ontologies. Much work in organisational theory has studied innovation in material terms and only recently has work begun to focus on relational meaning structures (Kirchner & Mohr, 2010). Work in economic sociology has begun to consider the cognitive embeddedness of producers in the market (Kennedy, 2008), but not other aspects of the product ontology. While this is in line with a field conceptions of a market (Fligstein, 1996; White, 1981), this neglects other forms that shape cognition around market information (Anand & Peterson, 2000).

Few studies have thus far looked at the emergence of product ontologies, nor have they considered what an empirical product ontology would look like. Most have been historical examinations of archives and patent records (Petroski, 1992; Friedel, 1994; Pinch & Bijker, 1987). Only a few studies have used media discourse to trace product ontologies through the nascent market stage. Rosa et al. (1999) studied the emergence of the minivan through a content analysis market stories published in the media and trade press. Kennedy (2008) studied just the one aspect of the product ontology with a producer network.

This thesis will explore the proposition that product ontologies may develop as meaning structures prior to a field configuring event. Because the field as a community may not be stable and the level ambiguity is extreme, it is not yet possible to study market information and stories. Therefore, a precondition for studying the initiation of product ontologies occurring in market information regimes is to look into rumour and speculation. Product publics are proposed as discursive arenas that will deliberate on rumour as a form of *improvised news* (Shibutani, 1966). This proposes that innovation rumours may initiate product ontologies whilst inventors and inventing organisations are still attempting to form them. The research question I will address is then, *how does ambiguity get resolved in the early moments of nascent markets?*

The proposed model in this chapter suggests that ambiguity is resolved in the early moments of nascent markets through a narrative that is told and retold by a product public. If such a narrative were to stabilise, this would form the basis of a product ontology. In order to explore whether this has empirical support, the investigation will explore two sub questions. First, *are product publics more likely to be found on the front stage or backstage of online discussion forums?* Second, if it is possible to sample product publics, *how will a meaning structure stabilise relative to a field configuring event?* The next chapter will outline the methodology used to answer this question in my empirical study.

Chapter 3 - Methodology

3.1 Introduction

The research question I developed in the previous chapter was, *how does ambiguity get resolved in the early moments of nascent markets?* To address this, I proposed several dynamics. The first suggested that early market discourse develops within an emergent product public. I developed this as part of a preliminary theory around *product publics* being temporarily constituted interaction-contexts that deliberate and develop discourses around emerging products. As loosely tied collectives, they are not communities with well formed social structures, but are instead discursive spaces made possible through mass-communication. Such publics are proposed to come about within social media. However, this leads to the issue of trying to locate product public activity in online spaces. The second dynamic was that of a product public developing a product ontology as a meaning structure. Unlike market stories, which are generated by legitimate market intermediaries, a *product ontology* develops within a *product public* in the early moments.

To explore this the research question empirically, I break the problem down into two smaller questions. First, *are product publics more likely to be found on the front stage (ie. Homepage) or backstage of online discussion forums?* This is a question that contrasts public-like and community-like activity as opposing organising principles. One is based on temporary interaction and attraction to specific topics of interest, and the other is based on more stable social ties, where interaction should

cut across topics. Second, if it is possible to sample product publics, *will a meaning structure stabilise in advance of a field configuring event, or will this just be noise?*

The previous chapter suggested that a *product public* develops a product ontology as an emerging discourse. Therefore, investigating this phenomenon will require a discourse analysis. I suggest here that a computational discourse analysis will enable both the search for public-like activity and the tools to trace meaning structures. This chapter will outline the methodology used to answer this question in my empirical study.

A site to study this phenomena should contain sufficient discussion around an emerging product. This will occur in the early moments, which I define as being in advance of a field configuring event (Meyer et al., 2005). Therefore I can use the event as the end point in the early moments, or pre-history. The criteria used to select a site will be: *a sufficient volume of speculation and rumour discussion, occurring in social media, and about a product that breaks out into a new category.* I propose that tech blogs discussing the tablet fulfil the criteria for an empirical site. Apple and Microsoft both announced tablets in January 2010, and shortly after, the tablet stabilised into a cognate product category residing between smartphones and laptop computers. Tech blogs have an accessible paper trail and often contain both a curated homepage written by an editorial team, as well as comment sections and forums written by users. Posts and discussions containing the word *tablet* continue well back to 2006. This leaves almost a five-year period of deliberation not affected by any official legitimate signals of intention from eventual market producers such as Apple, Samsung, Google, and Microsoft. Thus, the case of a tablet is compelling to

investigate a product public being enacted on tech blogs while deliberation an emerging product ontology occurred.

For this study, I collected every post published from 2006-2010 from four tech blogs that each span a range of perspectives. These include two company-focused sites and two general tech sites. I then constructed a relational database to structure the mined data and enable exploratory analyses. In such analyses, it was not clear whether product publics would be found on home page articles or in the comments and forums. This prompted another round of data collection, gathering a complete snapshot of one site for a full 13 months. I then analysed the structural dynamics of posting activity to find evidence of product publics and community activity. Once I was able to determine where product public activity was taking place, I conducted a discourse analysis based on computational linguistics to search for emerging meaning structures.

3.2 Studying early moments

3.2.1 Studying early moments of a product market

In selecting a site to study the early moments of a product market, there are significant questions of where to look and what to look for. In developing a discursive view of product market emergence, Chapter 2 proposed that the early moments are characterised by emerging discourses around product meaning. This suggests that a discourse as a product ontology will cohere and stabilise sufficiently enough to track it. Following such a structured discourse is linked to finding a product public. In choosing a site, I identify three criteria:

1. Sufficient volume of speculation and rumour discussion,

The first criterion for studying the early moments of a market from a discursive view is to find a site that is sufficiently engaged in discussion surrounding the upcoming product. In the early moments, these discussions are in the realm of rumour and speculation, as no official announcements, or field configuring events will have occurred yet. This is a pragmatist view of rumour being “uncertain knowledge”, where it has not yet been verified as being true or false (Fine & DiFonzo, 2011).

2. Occurring in social media

Publics are made possible by mass media, as they bring together actors from across social contexts. Social media is a particularly relevant type of mass media where deliberative space enables multi-way information flow. Moreover, the accessibility and reach of the Internet cuts across social and organisational contexts. With minimal transaction costs, information in social media can travel across firms, user

communities, and other institutional actors. This makes social media an ideal place for product publics to emerge. Furthermore, social media is occurring in electronic texts, so this corpus can be mined and studied.

3. About a product that breaks out into a new category

A necessary condition for studying the early moments of a product market is to find a case where a new market is actually emerging. The difficulty here is in finding a recent and large enough product category where its prehistory can be studied effectively. Social media has only been around for a decade, so this further complicates the selection of a product category that has emerged in this timeframe.

3.2.2 Studying product publics

The first step in this study was to find evidence of product public activity. The *product public* concept was developed in Chapter 2 as a proposed site for emerging discourse in early moments of a nascent market. Unlike the community structure based on reputation and identity, a public is a loose gathering of actors suspending their identities and engaging in deliberation. The public allows us to consider non-binding forms of deliberation, unimpeded by institutions and legitimacy. Temporary suspension of identity can allow for new forms of concerted action to emerge (Mische, 2009). Publics are made possible by mass media, and social media is one such channel where this may be occurring.

Most work in online spaces is based around the organising principle of stable social ties. When social ties are considered from the structural level, then community is a relevant organisational form. In a community structure, there are dense strong ties constituting a core, and weak ties extending into a periphery. This can be characterised by a power law of activity, whereby the majority of activity comes from the core users. When I relax the assumption of interaction as coming from social ties, then several alternatives can help to explain the mechanisms of online activity. One such mechanism of interaction is the relation of an actor to a topic. This alternative proposes the dynamic is based on the relation of *actor -> topic*, rather than *actor -> actor*. Evidence of such activity would support the organising principle of a public.

Publics have been theorised since Tarde (1898) and Dewey (1927) as coming about from mass media. This is related to the *public sphere* (Habermas, 1992), but has notable differences that surround heterogeneity. In considering the empirical form of a public, there are several defining features that include: *ephemeral interaction*, *temporary suspension of identities*, and *topical focus* (Mische, 2009; Shibutani, 1966; Dewey, 1927). Shibutani (1966) explicitly made the distinction between a public and a community,

“[the public] not an organised group, except in those instances when a crisis happens to be confined to an already established unit. It does not have a fixed and easily identifiable membership, nor is there a formal organisation of conventionally defined roles...Furthermore, a public is not to be confused with a community, for publics do not necessarily consist of the inhabitants of a particular territorial unit.” (Shibutani, 1966: 38)

Offline, the community may be linked by an identity based on a geographic territory and local proximity. Online, it becomes difficult to make this distinction, though perhaps territory could be considered in the sense of claiming jurisdiction over a

conceptual space. Members of an online community may feel a shared “consciousness of kind” (Muniz et al., 2006), as they form a sense of identity after spending time communicating together. Despite never meeting up in person (Faraj et al., 2011), online users may form a shared identity over a brand narrative as part of a “brand community” (Muniz et al., 2006).

The marketing literature has developed a body of work around online brand communities (Muniz & O’Guinn, 2001). By using “netnography” (Kozinets, 2002) researchers can find that tribes of consumer adherents celebrate brands through retelling of the historical narrative (Muniz et al., 2006). A pertinent example is that of the “Apple Macintosh Brand Community” that is described as almost religious (Muniz et al., 2006) and is often titled the “Cult of Mac” (Kahney, 2006). This community is depicted as being homogenous and mostly made up of “hard-core” passionate members who attempt to differentiate “true members” from not. Muniz et al., (2006) explored the role of rumour in the Apple brand community and suggested it was a common practice. Therefore, key assumptions about the Apple brand community include homogeneity, an intense passion, and sustained interaction over time.

Faraj et al., (2011) proposed a dynamic view of online communities where passion, time, and user interaction are resources that dynamically flow in and out of each community. When this view is contrasted with the brand community, there are some clear contradictions. Therefore, we can potentially recast the Faraj et al. (2011) view as being about online publics. When we relax the assumption of homogeneity and think about online activity as the flowing of resources in and out of the site, then we can think about community and public as being organising principles. Such an

analysis is contrasted to hard-bounded empirical constructs. In the case of online Apple adherents, there could be community-like activity and public-like occurring on the same platform.

In the organising dynamic of a product public, some users may have little passion or dedication to the brand, but come to deliberate and develop rumours as unverified propositions about a future state or product. Publics are linked to rumours in that every rumour has one (Allport & Postman, 1947; Shibutani, 1966). In most situations, publics form as temporary arrangements of social actors who deliberate a rumour as a collective sense-making process. In such a public, the bonds are not quite strong enough to support a shared sense of identity or kinship network.

Online communities have been studied fairly extensively by the innovation studies discipline, particularly with user innovation and open source software. This has focused on a combination of structural dimensions and individual motivations for participation (Lerner & Tirole, 2002; Larkhani & von Hippel, 2003; David & Shapiro, 2008), community governance (Murray & O'Mahony, 2007), boundaries (O'Mahony & Bechky, 2008) and roles (Nambisan & Baron, 2010). As online communities are relatively new social entities, it is not surprising to find such a framing. It has been just over a decade since Rheingold (2000) first made the case for the utopia of "virtual communities." Faraj & Johnson (2010) make the point that certain patterns of network behaviour reinforce the maturity and longevity of such communities. Indeed sites such as Slashdot have developed strong social norms (Gonzalez-Bailon et al., 2010). Yet our understanding remains surprisingly structural and static. As evidenced by work in online social network analysis, the focus has been on static structures of

embeddedness, with flow as probabilistic based on properties such as centrality (Borgatti & Foster, 2003; Burt, 2004).

There is a useful distinction to be made between the front stage and backstage of online communities. As the outcome of knowledge collaboration activities is shown to the world as a front stage performance (Goffman, 1959), it is accessible. Conversely, the backstage is like a community-based enactment that is messy, complicated and not easily accessible to newcomers or outsiders (Faraj et al., 2011).

Considering the scenario where people are interested in a topic, perhaps due to an unrealised market need. This topical discussion may take on the character of rumour, as conjecture and speculation emerge. Users who are interested may start to get involved, despite not necessarily having connections to other users. Over repeated posting they may form relations as social bonds. However, these relations remain weak as these users choose to post only in some threads pertaining to the topic. The speculated new product concept gets debated, developed and once the product is released, the same people do not come together in the same discursive combination. This activity of people coming together to debate a specific issue would be an online product public.

The work in online discussion networks considers how the flow of information is linked the structure of the network. Much focus is given to the mechanisms that explain these structures. Users enact roles such as the “answer person” that show up as structural signatures in the discussions (Welser et al., 2007). This leads to the question of whether users adjust their behaviours according to the expectations of a

context (Fisher et al., 2006). If this were indeed the case, roles would aggregate into patterned network structures. An alternative explanation is that users self select in topics that correspond to their expertise. In this opposing view, variance in discussion network structures comes from different compositions of users. The results from Kaltenbrunner et al., (2009: 4) show that,

“users adapt their patterns of interaction according to the setting in which their interactions take place, in this case determined by the topic of the discussion”.

This shows that topics are social contexts that adhere to norms and expectations, and are not simply the outcome of the aggregate of characteristics from self-selected users. From these results, we can start to think about public and community-like activity as being mechanisms that lead to particular network structures. This suggests that different structures are likely to appear according to topics of discussion, and this leads to:

Proposition 1: public and community activity are organising principles that create different shapes of discussion network structure

3.2.2 Studying emerging product ontologies as relational structures

In order to study a product ontology as a latent structure in texts, several criteria must be considered. First is the question of what counts as data in inferring the cognitive structure. Although market stories are likely to appear once there is a stable idea of the field actors, early market information can be found in gossip and rumour (Anand & Peterson, 2000). Therefore, an emerging product ontology in early moments will come from talk of speculation and rumour around a product concept. Second is the question of where to sample this data. Again, in the early moments of a product market, the network of market actors is not fully formed. Then a community-type frame may be inadequate for tracing early market information in a field. The *product public* allows us to conceive of a collective that is sufficiently engaged in deliberation, but not densely connected enough to meaningfully discuss social bonds. These two conditions can help to constitute the criteria for studying a nascent product ontology. Third, a product ontology can be empirically studied as a meaning structure that is derived from both early stage speculation discourse, and from mature stage market information. The product ontology concept has not been empirically developed. It has a strong theoretical precedent (Porac et al., 2002; Rosa et al., 1999), but it is based on a narrow set of historical studies. In operationalising it empirically, I consider it as a relational structure of concepts.

A relational meaning structure (Mohr, 1998) is an ordered set of concepts that are linked in a meaningful arrangement. Porac et al., (2002) defined the product ontology as a shared mental model that contained an array of attributes and other cognitive associations. This definition was implied to be relational as some concepts were

more embedded than others. For instance, the label of a stable product ontology will be the most central in invoking meaning. Therefore, thinking of a product ontology as a relational meaning structure is to consider it in concept network terms, similar to a web of meaning (Geertz, 1973).

Related work on social ontologies has also emphasised relational properties. Ruef (1999) developed a relational view of social ontologies to study changes in discourse in the US healthcare field over time. This was based on ideas laid out by Wittgenstein (1958) on meaning-in-use as an intersubjective achievement imperfectly overlapped as *family resemblances*,

“If you look at them [the category of games] you will not see something that is common to all, but similarities, relationships, and a whole series of them at all... We see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail. I can think of no better expression to characterize these similarities than ‘family resemblances’ ” Wittgenstein (1958: 31-32)

This perspective looks at connections between symbols as being links in discourse. Ruef used a scaling technique and co-occurrences of symbols in text that was similar to Carley’s (1993) map analysis, and a linguistics inspired “collocation analysis” (Pollach, 2012). Although he developed this as a relational theory of meaning, he did not project the symbols as a network, but rather used a weighted multidimensional scaling technique based on distance measures.

Several other recent papers in organisational theory (Kennedy, 2008; Tsoukas, 2009) have been attempting to project discourse in network form, but only Bingham & Kahl, (2012) have directly studied the text to find evidence of cognitive schemas. Kaplan (2011) described how “schemas”, “technology frames” and “cognitive maps” were

being used synonymously in the literature. She also outlined how recent advances in computational text analysis should open up new avenues for studying them. I propose that building from Carley (1993) and Corman et al. (2002), it is possible to analyse an empirical product ontology through computational linguistics tools and extend Ruef's (1999) relational view.

To explore a relational view of a product ontology, a method must be able to find the core conceptual links that make up a nomenclature. In a cognitive relational projection of an ontology, Ruef (1999) draws upon the idea of overlapping concepts to consider "family-resemblances". This is to say that meaning-in-use is not necessarily strictly bounded. Ostensibly, as an ontology is stabilised and reinforced within an epistemological domain, this web of meaning (Geertz, 1973) will likely take on a rigid relational structural form. However, a meaning-in-use in less formal contexts will inevitably be fuzzy and loosely coupled, particularly as it emerges.

As applied to an empirical product ontology, then the analytical strategy here would be to link common-sense understanding of a product market to a common nomenclature used to distinguish one category from another (Porac et al., 2002). In contrast to a dominant-design view of a product market that focuses on specifics such as material form and performance metrics, the product ontology is about a nomenclature as a family resemblance. Thus this allows us to consider multiple products existing within the same product ontology despite having differences.

The key brief of any study on social ontology is to "find the cognitive rules that are used to partition types of social organization" (Ruef, 1999). Determining the product

ontology is an interorganisational sensemaking process that results in the social construction of a product concept. To discuss the emergence of a product ontology, one must consider the questions of when the concept network becomes stable and the ordering of how its elements become cognitively embedded. Kennedy (2008) has demonstrated the embeddedness of a producer network, but Porac et al. (2002) describe other elements such as a constellation of attributes, use-cases of the product and an embeddedness of other market actors beyond simply the network of producers. These elements each represent different aspects of the product ontology and an understanding of emergence would need to study their ordering and centrality in the cognitive embeddedness.

3.3 The tablet computer and tech blogs

In this section, I use the criteria developed previously to describe how the case of the tablet discussed on tech blogs is an appropriate site for studying the phenomena of a product ontology emerging in the early moments within a product public. In the first part of this section, I describe how the case of the tablet computer product is an example of how the pre-history of a market helps to determine its emergence. In the second part, I describe how this pre-history can be studied through tech blogs.

3.3.1 The tablet computer

There was exceptional media focus around Apple's iPad product as a technology drama (David, 1992) in its unveiling. This was an innovation disclosure event where Steve Jobs casually sat on a couch on stage and held up to the world his "magical" product. This keynote speech at Macworld San Francisco kicked off a field configuring event (Meyer et al., 2005), where Apple, the media and other technology developers came together for the unveiling of the much anticipated apple tablet. This marked an epoch in the emergence of the tablet market. Within 60 days from launch, several competitors began to crowd this space as Apple's category straddling reached two million sales. It would be reasonable to suggest that the emergence of the market following this epochal event was remarkable in speed and scale. However, the pre-history reveals that many of the conceptual foundations had been established in discourses beforehand.

At the time of this writing, the tablet is now an established mainstream product, with 2013 estimates suggesting unit sales of 121 million (NPD DisplaySearch, 2013). Further reports are also suggesting that laptop PC sales are dropping and the tablet market will surpass it by 2016. This level of market disruption is tremendous and seems to be occurring at a rapid rate. Another indication of the tablet market emerging in 2010 was that market analysts such as the NPD Group began compiling worldwide sales figures as separate from laptop PCs. In less than 6 months after the iPad's release, the space was crowded by announcements from competitors such as HP, Microsoft, Dell, Asus, Samsung, Google, RIM, and HTC (Dybwad, 2010). According to the technology life cycle model of market emergence (Geroski, 2003), the era of ferment should require adequate time for variation, selection and retention towards dominant design. However, in this market, it appears dominant design was established rather quickly around the Apple iPad. Even by liberal assumptions in the model, this was a rapid stabilisation. This leads us to consider the origins of the market before Apple's innovation disclosure event.

In 2001 Microsoft began promoting a product they called the "Tablet PC". Bill Gates showed prototypes of it around to industry events and framed it as an evolutionary step for the laptop PC. It was to be the next step in computing interfaces, taking advantage of digital ink and handwriting input (Lohr, 2002). Gates predicted that within five years, the Tablet PC would become the "most popular form of PC sold in America" (Microsoft, 2001). However, by 2002, when partners began producing these machines, it was clear that Gates' ambitions had been scaled back. The eventual product remained a laptop, and was enhanced with pen-based features. Despite Microsoft's efforts and those of major computer makers Acer, Fujitsu, HP,

Toshiba, NEC to push these hybrid machines, the market category never really took off as a mainstream product. By 2005, the Tablet PC had accounted for 1.5 percent of laptop PC sales, translating to one million units shipped worldwide in that year (NPD Display Search, 2013).

In diagnosing the failure of the Tablet PC to become a mainstream product, one could consider a multitude of factors. Perhaps the technology presented was not mature enough, or perhaps the concept was not distinguished enough from the laptop PC and understandings were simply subsumed back into that category. As a disruptive innovation, the Tablet PC was a fringe product, so long as the dominant frame was the laptop PC. However, this does speak to the germination period of the tablet computing concept.

In the 2012 patent infringement lawsuit between Apple and Samsung around tablet designs, lawyers on both sides presented evidence as to the history of the tablet concept and their company's claims to it. Samsung's legal team presented a case that included historical prototypes from assorted companies and movies. They presented prior-art of tablet computing concepts found in the science-fiction film, *2001: A Space Odyssey* (Clark, 2011). In turn, Apple's lawyers presented evidence claiming an evolutionary lineage from the Newton to the iPhone. Apple eventually won the case, but it was clear that the concept of a tablet-like computer had been around for at least four decades before Steve Jobs' dramatic introduction in 2010.

Around Jobs' keynote address on January 27, 2010, the world seemed to be ready for the "post-PC" tablet, anticipating what he would introduce, with specific detail. By

2009, the eBook readers market was growing and the iPhone was a phenomenon (Stone, 2009). By then, the mainstream media occasionally published speculation that tracked concepts of what shape the Apple tablet might take. The New York Times gathered speculation from tech blogs and published stories asking, “What Would a 10-Inch Apple ‘iSlate’ Look Like?” (Bilton, 2009), and “Would You Buy the Mythical Apple Tablet?” (Stone, 2009). More intriguing is how Steve Ballmer, the CEO of Microsoft, unveiled a tablet-like product that was remarkably similar to the iPad two weeks before Jobs’ went up on stage (Vance, 2010). From this we can say that several companies supposedly working in isolation had come up with vastly similar designs and announced them to the world within a span of two weeks. Somehow information was able to flow in a space that allowed the seeds of the tablet market to grow.

3.3.2 Tech blogs

In recent years, a form of website has emerged where the site is like a hub dedicated to exchanging informal information about innovation in firms and industries. One example is that of Apple Computer (Stelter, 2008). Muniz et al., (2006) described this phenomenon as, “Rumors are such a common part of the Macintosh community that there are multiple websites that specialise in debating and analysing Macintosh rumors” (2006:241). A look into activity across “rumour websites” such as *www.macrumors.com* reveals a keen interest in a company’s products, services and brand. The dominant use of such websites seems to be in trading rumour and speculation regarding future unannounced innovations. Included in such discussions

are spy-photos and mock-ups of prototypes, debates over feature lists, and discussions of price and release dates.

“DUNN: Apple's cult of secrecy has bred a whole industry of bloggers like [John] Gruber. He describes himself as the Roger Ebert of the industry - mostly a critic of Apple products. But some of his peers pride themselves on their investigative reporting. Neil Hughes works for AppleInsider with a team of eight other reporters in Punta Gorda, Florida. They compete with sites like MacRumors, Boy-Genius and MacDailyNews to break stories about Apple products.” (NPR broadcast, December 7, 2009)

As the NPR broadcast suggests, many of these sites are tech blogs that cover Apple closely.

The case of rumours concerning Apple products provides one example from consumer electronics. Other examples of consumer electronics product publics could well be found for mobile phones, photographic equipment, and audio equipment. For Apple products there are over ten websites that can be noted as hosting “rumour sites”, but the most prominent is macrumors.com (Stelter, 2008). From a rumour standpoint, the members who post to this website self-select as being sufficiently interested and are perhaps affected by the outcome of Apple products. It is not entirely clear why they would seek information about current and forthcoming products. One would expect that other members, who may or may not include those posting material, would be Apple employees, competitors to Apple products, stock analysts and those who make complementary goods. In addition, there may be those who are looking to analyse potential activity of the market, such as industry analysts and journalists.

Recent work has started to look at technology blogs (Davidson & Vaast, 2009; Droge et al., 2010) as constituting a community of core-periphery connections that are potentially involved in framing market information. It is not yet clear how tech blogs are involved with the naming and categorisation work, especially as less recognised actors in knowledge institutions of information technology (Pollock & Williams, 2011). Davidson & Vaast (2009), observe how technology innovation discourses were developing in online communities of 'tech bloggers'. Brought together by common interests and behaviours, the tech bloggers are described as "bloggers whose postings focus on developments and innovation with new technologies, particularly information technologies, and with high-tech business" (Davidson & Vaast, 2009). Here, they represent the core of a community as "A-lister" actors with influence in emerging discourses. These editors of tech-blogs are shown to be playing a role and gaining influence in the discourse community as,

"established institutional actors (corporations and MSM [mainstream media]) in the field of technology innovation discourse appear to be adapting to the emergence of a new actor (tech bloggers). In turn, tech bloggers establish their position and legitimacy through their discourse about their interactions with corporations and MSM" (Davidson & Vaast, 2009: 56)

In their exploratory study, Davidson & Vaast (2009) concluded that much work still needs to be done to explore how contributions to discussion and characterisations of technologies affect the actions of technology firms and potential customers.

Tech blogs were studied again by Droge et al. (2010) using an exploratory content analysis looking for whether blogs could provide input to new product development (NPD) processes similar to the way customers and suppliers could be "lead users" (von Hippel, 2005). Results showed that the tech blogs provided useful

information for new product development managers such as how and when the product would be used, some favourable and unfavourable ratings of trends, companies and products. The authors found that the blogger community could show confusion or ignorance about features or functions. They suggested that companies should be aware that tech blogs could potentially set expectations and the marketing agenda for a product market.

One aspect not covered by both Davidson & Vaast (2009) and Droge et al., (2010) was discourse emerging before products are launched. At this point in the product ontology evolution, because the company has not officially announced the product, any discussion around it will be in the realm of speculation and rumour. Recall the conception of rumours as uncertain knowledge lacking a reliable sponsor (Fine & DiFonzo, 2011). Pre-release product concepts in rumour are particularly interesting for examining the emergence of product ontologies.

Tech blogs are good candidates for constituting a space enabling product public-like activity. As forms in social media, they have less historical reputation and social capital than mainstream media. They have also been shown to engage in early discussions of product markets. Existing work on tech blogs has attempted to frame the collective structures and positioning of legitimacy in blog-communities (Davidson & Vaast, 2009), but there is also an acknowledgement that they may engage in early product talk (Droge et al., 2010). Therefore, tech blogs are good candidates for this study because they fulfil the criteria of sufficiently discussing product speculation, in social media, and are active during the lead up to the new tablet product market.

3.4 Data collection

3.4.1 Introduction

I collected two types of data for this study and processed them to generate a third. I data-mined both a complete snapshot from one tech blog of Jan 2006 to Jan 2007, and all homepage posts from four tech blogs for dates ranging 2006-2010. I used the former to find evidence of product public activity in the front page articles of the single tech blog. From there, I was confident to pursue using the front page posts on the four tech blogs as a site to explore the tablet. I sampled from the set of posts for those that contained the keyword “tablet.” I constructed a corpus from the sample, and organised the posts in ascending order by date published. Then by going through each quarterly set, I constructed concept networks.

3.4.2 Sampling a web forum to find a public: Macrumors.com

In this section, I explain the methods used to crawl to gather data for a complete snapshot of a tech blog over a full year. Following the empirical strategies of Gonzalez-Bailon et al., (2010) and Kaltenbrunner et al., (2009), I collected observational data of thousands of discussions taking place in their natural setting. The analysis aims to uncover structural signatures of communities and publics from discussion dynamics, by looking at the variables of user activity and post timing, against the type of thread.

The site selected was *macrumors.com*, which purports to be the site containing “news and rumors that you care about.” As a website, it is divided up into a front-page blog and a discussion forum. It is primarily an Apple-focused site and in 2011 was regularly getting more than 8,865,115 visits per month, globally (Quantcast, 2011). The founder Arnold Kim, who has been profiled by the New York Times as a former medical physician, said the site was intended to be “an aggregator of all the [Apple] rumors and hints that appeared on message boards and other Web sites” (Stelter, 2008). The blog is maintained by a team of editors and the forum is open to all. It makes money through advertising on all of its pages.

The posts and threads are all publicly available to read, however, in order to post, users must register with a valid email address and unique username. From an initial analysis, I found the forum to be based on a modified version standard serving software called vBulletin. This is a popular piece of web-forum software has been studied by organisation scholars analysing other online communities (Faraj & Johnson, 2010).

The organisation of the web-forum is such that posts are contained with threads that resemble an ongoing communication, similar to an email thread. Each thread contains a description and is assigned to a category from a predetermined set maintained by site administrators. The thread structure is linear in that users post in a flat sequence. At the time of collection, there were no features such as voting on posts, as has been reported in studies on other online communities such as the site Slashdot (Gonzalez-Bailon et al., 2010). There was loose moderation by

administrators, where users started and maintained their own threads and could be removed from the site based on improper activity.

Macrumors.com is a website that is dedicated to talk of Apple products and brand.

Within most calendar years, Apple has several Innovation Disclosure Events, where new products are announced and released. These appear on an annual schedule with approximately four per year. An adequate sample to capture user activity would have to cover one full cycle to maximise chances of capturing appropriate activity around innovation disclosure events. Therefore, to get a sufficient snapshot of cyclical activity 13 months of data were crawled from the macrumors site. This date range is in line with previous online community studies, with the logic of a calendar year plus slightly more (Kaltenbrunner et al., 2009; Gonzalez-Bailon et al., 2010).

Within one full turn of the schedule, the dataset should contain sufficient evidence of the location of public-like activity and community-like activity.

Variables	Description
<i>Threads (N=133,555)</i>	
topic_id	Unique integer identifier of Thread
title	Title of the Thread
category	Each thread is fit into a hierarchy of categories
<i>Posts (N=1,460,102)</i>	
id	Unique integer identifier of Post
timestamp	Date and time (Month, Day, Hour, Minute)
body	Text content of Post
position in thread	Sequential order of post within thread
topic_id	Unique integer identifier of Thread post appears in
author_id	Unique integer identifier of Post author
<i>Commentators (users) (N=78,785)</i>	
user id	Unique integer identifier of author
userName	Text name
tagline	Text description of user
joinDate	Date user signed up

Table 3.1 Information crawled from the Macrumors.com Jan 1 2007-Jan 31, 2008

Table 3.1 describes the structure of the data crawled. Threads, Posts, Users are the main data entities, each with several properties and relationships across. Each thread contains a set of posts. Every post has an author and belongs to a thread. The data collected was subject to a series of verification checks to ensure that the entire data range from Jan 1, 2007- Jan 31, 2008 was collected. These included scanning the list of post-ids to ensure that all with a range were being collected. The dataset includes 133,555 threads, which contain 1,460,102 posts, written by 78,785 unique users. The thread-lengths are calculated as number of posts contained.

3.4.3 Sampling from a product public around a tablet: tech blogs

In this section, I describe how I sampled from several tech blogs that are proposed to participate in a product public around the tablet. For this study, I chose four blogs that reported for many years on the emergence of the tablet product category. The first, *macrumors.com*, was described earlier in the chapter. The second site chosen was *Engadget*, and selected for the wider audience of technology enthusiasts who frequent social media. *Engadget* is owned by AOL media, and represents an established technology blog. It was included the Droge et al., (2010) study, and was listed in 2010 as one of TIME magazine's best blogs (Fletcher, 2010). The third blog selected was *BoyGeniusReport*. Originally as an offshoot from Engadget, it was selected by aggregator Technorati as fourth highest "technology authority." The fourth site selected for sampling was *Crackberry*, a site dedicated to reporting on Research in Motion and Blackberry products. It was selected as a site covering a company that has a similar profile to Apple in the smartphone and tablet markets. If a

product ontology emerges in the industry, the nomenclature should cover Apple and its potential competitors. Table 3.2 lists details about each site.

<i>Dimension</i>	<i>Macrumors.com</i>	<i>Engadget.com</i>	<i>BoyGeniusReport.com</i>	<i>Crackberry.com</i>
<i>Foundation date</i>	29/02/2001	1/3/2004	1/10/2006	1/2/2007
<i>Founder</i>	Arnold Kim (former physician)	Peter Rojas (formerly Gizmodo)	Jonathon Geller (formerly Engadget)	Kevin Michaluk (former Journalist)
<i>Blog articles</i>	13,989	58,532	13,105	9082
<i>User feedback</i>	Article commenting & Forums	Article Commenting	Article Commenting	Article commenting & Forums
<i># of mentions of "tablet"</i>	605	3374	861	254
<i>mentions as % of all articles on site</i>	4.30%	5.80%	6.60%	2.80%

Table 3.2: Tech Blog descriptions

This data-mining process here for gathering stories from tech blogs is broken up into three phases: first, I constructed a text-corpus from technology blog articles through data-mining. Following guidelines of ethical blog usage in social science (Olson, 2008), I selected websites that offered up public access to data. This then made my actions analogous to a search engine, indexing pages on their site. For *macrumors.com*, each article appeared as a link in an index page with a listing such as "Archives - January 2001." On such a page, each article link appeared as similar snippets of HTML code. By capturing each month's archive page, I was able to gather links covering the complete set of articles. Collection was done through a custom written PHP script employing the cURL engine to gather pages and regular expressions to parse out patterned text snippets that matched the appropriate fields. This data was then stored in a relational database using MySQL.

Similarly, for data-mining of *Engadget*, *Boygeniusreport* and *Crackberry*, archive

pages were generated by manipulating a URL, incrementing a main page count until the first page was retrieved. An example is: <http://www.engadget.com/page/2/>, where “2” indicates it is the second page back in time from the current date.

Each post was collected, and the text cleaned and stored in the relational database. From this database, the dataset could be resampled according to interesting parameters such as time, site, and keywords. From my ethnographic exposure to reading thousands of texts, it was clear that the keyword **tablet** was an anchor for discussion. By incorporating the wildcard, variations such as mactablet were included. Other potential keywords such as slate, iPad and ebook resulted in a similar dataset to the query for **tablet**.

3.4.4 Projecting concept networks from text

In this section I describe how I resampled posts from the database according to longitudinal slices by yearly quarter and constructed concept networks. In order to perform this analysis, I needed to develop a novel method for constructing concept networks from text. The analysis for this study entailed using each constructed concept network as a projection of any potential meaning structures from within the corpus of text from each quarter. As a longitudinal study, I compared each quarterly concept network by using the structural properties of each. I will outline how this method works.

Carley (1993, 1997) developed a methodology entitled “map analysis” as a relational perspective on concepts and meaning. Her critique of classic content analysis

focused on the lack of a relational dimension when counting concepts “in isolation.” Map analysis was an attempt to bridge the contributions of content analysis with a relational view. As an organisational sociologist, she applied these efforts towards social cognition and finding means of mapping consensus. Map analysis was based on the idea of a “conceptual network.” That is, a set of concepts and the relations between them made up network ties. Given that language could be represented as a network of concepts (Axelrod, 1976), map analysis offered up a method for operationalising social knowledge.

Carley (2002) called for organisational theory and sociology to move beyond the content analysis approach of counting concept frequencies in isolation and to develop relational methods that integrate text analysis with network methods. This coincided with the emerging discipline of relational sociology that aims to move beyond social network analysis and to integrate discourse and culture (Kirchner & Mohr, 2010). This also was initiated somewhat by White’s (2002) recognition that culture and publics as discursive performances were important for analysing social structure. Others have attempted to embrace networks and text analysis, often under the frame of a cultural approach to organisations (Mohr, 1998; Pachucki & Breiger, 2010; Weber & Dacin, 2012).

In the strategy literature, Kennedy (2005, 2008) projected text-networks based on producer co-mentions in media articles around a nascent market. In management, some recent attempts have been made to embrace corpus linguistics as a means to extend text analysis (Pollach, 2012). Kaplan (2011) describes a trend in strategy and

cognition of textual analysis of themes and relationships. She suggests it is being driven by availability of texts and sophistication in computational methods.

Weber & Dacin (2012) describe the shift in cultural research in organisational studies and the conceptual vocabulary, treating culture as a dynamic “agentic way.” This can take the form of applying network analytic models to analyse cultural artefacts such as texts, conversations, tastes, and rituals. If culture is the Geertz’s “web of meaning,” then meaning is associative, and the shape of different webs can be mapped with standard network analytic routines (Breiger 2000, Mohr 1998). For content analysis methods, this is clearly moving beyond the simple word counts of the past.

Carley (1993a, 1993b, 1997b) has been a central figure in developing a semantic-network method for sociology and organisational studies. In order to appreciate meaning in texts, she argues that relationality of concepts should be taken into account. Concepts together form meaning through their patterns and this can be mapped out (1993). When concurrent texts from a social context are mapped out, this can provide the basis for extracting shared mental models (1997a), culture (1994) or social cognition (2002). One promise of this method is to help identify social consensus, as this analysis “may provide insight into the use of language by members of the society to build consensus” (Carley, 1997b).

Multiple techniques currently exist for extracting conceptual networks (Carley, 1986; Carley & Palmquist, 1992; Franzosi, 1994). The basic model from many of these is articulated in Carley’s conceptual network (1997). Her starting point builds from

Axelrod (1976), in representing language as a network of concepts and the relationships between them. This network can be thought of as “the social structure of language... or, the representation of extant social knowledge” (Carley, 1997).

Carley argues that each concept has a position in the network that can be characterised along several dimensions that include imageability, evokability, density, conductivity and intensity. These dimensions provide a relational framework to analyse concepts.

The base elements of Carley’s (1997) conceptual networks are concepts and relations between them. Here, a concept is a single idea, or ideational kernel. Without the relation to other concepts, it lacks meaning. The relationship between two concepts is important, encapsulating strength and directionality. In a socio-cognitive network, the strength of a relationship can indicate the degree of consensus. Directionality in a relationship refers to placement of concepts, with either in the anterior or posterior position. Together, two concepts and the relationship between them is known as a statement.

Here, the sentence is a basic unit of thought (Carley, 1993). Every sentence is made up linguistic elements and will include at least a subject and object. These linguistic roles are both part of noun phrases. Natural language processing research has focused intensively upon noun phrase extraction (Manning et al., 2008). Most of these techniques follow the workflow of first tagging words in a sentence according to their Part Of Speech, then parsing the grammar trees to find word-chunks such as Verb phrases and Noun phrases. These noun phrases are taken to be concepts (Manning & Schütz, 1999). Concepts strung together in a web represent a text.

Betweenness centrality (Freeman, 1977) is an important mechanism for extracting key noun phrases. This is based on the idea that Noun phrases appearing across different sentences effectively serve as broker nodes. A text is then held together through concepts that bridge sentences and create a structure. It is this latent structure that generates meaning.

3.5 Developing a novel method for projecting concept networks

3.5.1 Introduction

In this section, I outline the development of a novel method for projecting text networks. This builds from Carley's (1993) conceptual networks and the Centering Resonance Analysis (Corman et al., 2002) approach of using Computational Linguistics to extract noun phrases and finding sentence overlaps. This turns out to be a form of computational discourse analysis that combines Carley with meaning structures (Mohr, 1998). By extending these methods, I am able to build noun phrase networks that can help track a product ontology. This method differs from the aforementioned in three key aspects: *using multiword noun phrases instead of single word influence, using community detection modularity to measure cohesiveness, and using fuzzy-string comparison to construct the network*. The outputs of my method are threefold: *modularity scores, rank order list of central noun phrases, and network visualisations*. I step through each element of the method, then illustrate its usage with an example.

3.5.2 Concept networks

Historically in social science there has been much interest in meaning constituted through associations of symbols. These associations may even speak more to meaning than the symbols themselves, as De Saussure [1916] (1966) claimed that within relations “there are only differences.” Wittgenstein (1958) proposed a relational view of meaning with a loose network of similarities and relationships between concepts. In early semiotics, Pierce (Hodge & Kress, 1988) used “infinite semiosis” to describe how the definition of signs came from relations with *interpretants*. In 1909 he proposed a graphical notation of “existential graphs.” Extending these ideas to culture, Geertz (1973) also drew upon these ideas to argue that, “man is an animal suspended in webs of significance he himself has spun, I take culture to be those webs.” Geertz outlined the role of ethnographers to ‘inscribe’ social discourse, creating interpretive accounts of events through thick descriptions. This is an important contribution to any social theory of meaning, as actions such as gestures or tone may convey just as much by way of meaning as discourse. One way Organisational Theory has circumvented this problem is by focusing on legitimated institutional discourse in fields, where meaning is formalised by a network of actors. As an example, Lawrence & Phillips (2004) showed how the local actions of an institutional entrepreneur helped shift the web of meaning around whales from aggressive and dangerous to friendly endangered animals. This was done through multiple discourses as a new institutional field was created. Discourse analysis then has particular importance in mapping the emergence of meaning.

There has been recent interest in Organisational Theory of expanding computational-aided text analysis by borrowed methods from Computer Science and Linguistics (Pollach, 2012; Kaplan 2011; Ramage et al., 2009; Oard, 2009). With the

increased electronic availability of texts and sophistication in computational methods, this has opened up opportunities for analysis by enabling a larger scale empirical study of established theoretical constructs.

Carley's (1986, 1993, 1997) efforts have been critical in developing a methodology for measuring cognitive maps. Her work has built the basis for a family of techniques known as Mapping Analysis. In this approach situated concepts and their relationships are studied. This is based on the proposition that language can be represented as a "lossy-integrated conceptual network"¹. By definition, the conceptual network is a set of concepts and pairwise relations between them.

The map analysis method is a form of text analysis that differs in many respects from traditional content analysis (Krippendorff, 2003). In a typical content analysis, concepts are identified through coding dictionaries and are counted in relative isolation. Inferences about meaning are argued to come from frequency as one corpus is compared to another. This approach can answer comparative questions about word distribution or the proportion of types of words, but does have to draw upon an explicit theory of meaning. Carley (1993) suggested that map analysis was a broader category, which subsumes and extends content analysis techniques. By incorporating pattern-level analyses, it enables the mapping of concepts from textual discourse.

In the base form of Carley's method, maps are networks of concepts. The concept is an ideational kernel, a single idea devoid of meaning, except as it relates to other

¹ here, lossy refers to information that is discarded in the process of projecting a network

concepts. Relationships link concepts and help to comprise meaning in a structural arrangement. The common operationalisation of this method is based on isolating sentences as basic units of thought (Carley, 1993). Then concepts are identified from the sentence as nouns. The relationships between nouns are verbs, so the network tie is a directed one. The resultant network that is projected from this is *lossy*; that is, information is lost as the grammatical structure of the sentence is simplified and words are dropped. This allows the potentially similar nouns to overlap across sentences.

Carley's method aims to uncover the relationships of concepts through an intensive analysis of a small text corpus. This has been used to extract the structures of mental models from teams (Carley & Palmquist, 1992) and culture (Carley, 1994). At a social level, this cognitive mapping can be used to measure socio-cognitive schemas from written discourse (Barr, Stimpert & Huff, 1992; Tsoukas, 2009). Bingham & Kahl (2012), in their study tracing the emergence of the cognitive schema of the computer, analysed textual discourse from the insurance industry over 1945-1975. However, this study was largely qualitative, with a human coder reading the corpus and generating a list of nouns and verbs. This was coded again through a second reading of each text to find relations. The advantage of this manual approach is the ability to overcome linguistic ambiguities and code for a relatively complex relational structure. The downsides to this approach are scale, scope and complexity. The reliability of the coder may not be consistent, nor will they be able to manage large corpora.

A series of methods have emerged based on map analysis (Carley, 1993) that could be grouped together as “computational text-analysis.” These approaches also process sentences within a text to isolate concepts and their relational structure. Ruef (1999) used computational text analysis to build a relational view of social ontologies in the US Healthcare field. Kennedy (2005, 2008) used these methods to trace the emergence of cognitive structures in nascent markets. The advantages to a computational approach come from consistent reliability and by simplifying the fundamental structures that underlie the network projection. This enables more commensurability on larger scale of analysing text corpora. The limitations of a computational approach are many, particularly as the coding must be done through rule-based heuristics. This is particularly the case with language collected from informal settings, where formalised rules of grammar may not have been stringently applied. However, using machine-learning to parse grammar, known as “Natural Language Processing” (NLP), has been a major area of research activity in computational linguistics and much progress has been made in the last two decades (Manning et al., 2008; Manning & Schuetze, 1999).

Chomsky (1956) advanced the idea of nouns and verbs as phrases instead of singular words. This has been a difficult problem for automated methods, but there has been significant progress made in the accuracy of parsing grammatical structure (Manning et al., 2008). In Information Studies and Computer Science disciplines, a major focus has been on extracting noun phrases. There seems to be evidence that using noun phrases strung into networks holds promise for interesting applications such as identifying topics for extraction based summaries (Xie et al., 2004) or deriving conversational themes (Dwyer, 2012).

Phrase based structures are a powerful way to approach English grammar (Chomsky, 1956). This is clearly evident when considering proper nouns containing multiple words such as the names of people or organisations. In a text analysis, the words “New York Times” and “Wall Street Journal” are meaningfully grouped as functioning nouns. The Routledge Dictionary of Language and Linguistics defines the noun phrase as a, “grammatical category (or phrase) which normally contains a noun (fruit, happiness, Phil) as its head and which can be modified (specified) in many ways” (Bussmann, 1996). Thus, multi-term noun phrases can provide more detail than single term nouns. Extracting noun phrases has been a focus for Computational Linguistics and parsing methods for natural language processing are now sufficiently accurate for research applications (Manning et al., 2008).

Corman et al., (2002) introduced a computational method for network text analysis they called Centering Resonance Analysis (CRA). This method is able to project semantic networks from text corpora with the goal of finding influential words. The method is based on a theory of communicative coherence, which posited how texts are constructed around central utterances (Grosz et al., 1995). The CRA method has provoked minor interest in the organisational studies (Kuhn, 2008; Weber & Dacin, 2012; Grodal, 2007).

Centering Resonance Analysis is an automated computational text method that combines linguistics with natural language processing. It is based on two principal procedures. The first scans sentences sequentially and tags linguistic Parts of Speech (POS) tags according to each word. This information is used to find chunks

of words that make up noun phrases. By extracting the noun phrases from each sentence, the method is able to string together a chain of central utterances from the entire text. Each noun phrase utterance is evaluated according to its betweenness centrality (Freeman, 1977) in the projected semantic network and the head word is extracted. The second procedure is finding resonance between texts, where “the more two texts frequently use the same words in influential positions, the more word resonance they have” (Corman et al., 2002). The output of this method is a network of individual words that are evaluated as most influential.

The primary goal of CRA is to find influential words within a semantic network. Weber & Dacin (2012) suggest that methodological opportunities for text networks come as they are mapped with standard network analytic routines (Breiger 2000, Mohr 1998). Within social network analysis, the main two foci are node position and structural network properties (Kilduff et al., 2006). CRA is focused primarily on the former, looking for utterances, or noun phrase words, that are influential as broker concepts. A weakness of the CRA approach is that structural properties are not taken into account as an analytical tool. Furthermore, by flattening the noun phrase to a single word, the method is also at risk of breaking apart bigrams (such as noun phrases) that have limited compositionality as collocations (Manning & Schütze, 1999). Put differently, breaking apart words is at risk of vanquishing meaning.

The issue of with collocation compositionality is about unit level concepts as applied to noun phrases. An example of this is the bigram “strong tea.” Broken apart, this meaning is lost as each individual word corresponds to a separate concept. Manning & Schütze, (1999: 151) describe how in this example,

“strong has acquired the meaning rich in some active agent which is closely related, but slightly different from the basic sense having great physical strength.”

Extracting single words is important for normalising nodes of a network, but is at fault of breaking collocation meaning and is not necessarily representative of the bigram in which it frequently appears in the text.

The CRA method explicitly focuses two analytical strategies: on the placement of single words as nodes with high betweenness centrality within concept networks, and structural equivalence of those between networks. This can potentially allow the repurposing of the analytical ideas of social network analysis. However, this CRA work has remained marginal in social science.

Social network analysis methods have been criticised at times for being too focused on descriptive statistics of static network slices (Monge & Contractor, 2003). Watts (2004) suggested more could be done to focus on inferences of network dynamics and collective behaviour. P*/EGRM² methods (Contractor et al., 2006) are one such attempt to capture structural properties of the network in order to assist in inferential inquiry. Another is the emerging interest in community detection and modularity (Leicht & Newman, 2008). The CRA method is not a direct descendent of social network analysis, however the same critique could be levelled against it as an analytical tool. In my development of the novel method, I aim to address these critiques and opportunities.

² These methods are used to find statistically significant structural signatures by spawning large numbers of similar networks randomly and comparing to the focal network

3.5.3 Constructing concept networks

This section outlines a method for constructing concept networks. It seeks to address three gaps from previous approaches: *handling multiword noun phrases*, *collapsing lexically similar noun phrases into nodes*, and *network level measures*. The method developed here is an extension of both Carley (1993) and CRA (Corman et al., 2002) and has been implemented in customised software. I wrote the algorithms into a level of software code in Python on top of established open-source frameworks from computational linguistics. These frameworks include the Stanford NLP Parser (Klein & Manning, 2003); Pattern for Python (De Smedt & Daelemans, 2012); and the Natural Language Toolkit (Bird et al., 2009). Each of these toolkits have unique strengths and weaknesses, and due to their modular properties, are combined to implement the method outlined here.

The steps for constructing concept networks were explained by Carley (1993) as finding and extracting both noun phrases and verb phrases from texts. Within each sentence, noun phrases are linked in a directed manner by the verb phrase. A simple way to conceptualise this in linguistic units is to consider the subject verb object structure in a sentence. For the sample sentence, “*Steve ate the apple*”, *Steve* is the subject, *ate* is the verb, and *the apple* is the object. This can be depicted in the relation *Steve -> Apple*, where the arrow is the verb *ate*. Finding concept networks this way can be performed by a human coder or machine-learning algorithm.

The example of a simple sentence structure shows the most basic of sentences. In most texts, sentences will take on more complex structure. This makes the verb-

phrase dependencies more difficult to infer. Consider the slightly more complex sentence, “*Steve was an Apple picker who ate apples.*” This shifts the structure around, so that whilst *Steve* is still the subject, it relates to two verb phrases, *was* and *ate*. These verb-phrases then act on the two objects, *an Apple picker* and *apples*, respectively.

Grammars in some cases are ambiguously applied to well-formed sentences, but ambiguity is particularly salient when attempting to infer structure from *natural language* texts. In order to use Carley’s (1993) method of projecting concept networks as noun phrases acting on other noun phrases with each sentence, the linguistic units must first be isolated. Once these units have been extracted, the next step is to find relations between phrases. Pinker (2007) showed how grammar can be ambiguous along several dimensions, but this complexity is increased by the imperfections of writing in natural language. Beyond well-edited professionally produced texts, it is a safe empirical assumption that many texts will not adhere to a single formal grammar. Therefore, to apply the method to texts, grammar rules must be inferred in a probabilistic manner.

For the full implementation of Carley (1993), of relations between noun phrases, ambiguities of verb-phrase dependencies need to be resolved (Manning et al., 2008). If done by a human coder, it can be a cumbersome process and will limit the scale and scope of the exercise. If done by an undirected machine-learning algorithm, the complexities and ambiguities inherent in natural language discourse may be difficult to adequately resolve. The process of finding dependencies in grammars is known as *deep parsing* (Belz, 2002). This is a detailed approach that is

close to Carley's (1993) originally proposed method. However, given the complexities and ambiguities of deep parsing, an unsupervised machine learning method is potentially unreliable. This is an active area of research within the computational linguistics discipline and it is likely to become more reliable over time.

The *deep parsing* approach to map analysis focuses on the complex structures within sentences, but may also prohibit a scalable analysis across sentences. For the methodology this chapter aims to develop, the range of research questions it enables surround inferring meaning structures across texts. Therefore, I consider a scaled back approach to Carley's (1993) method that corresponds to *shallow parsing* (Belz, 2002). It is less complex, computationally more efficient, and still enables inference about structure. This approach to projecting text networks first tags words according to their parts of speech, and then groups them into meaningful phrases, or 'chunks.' These chunks can be noun phrases, verb phrases, or several others. Noun phrases are the primary focus, as they are the most common and serve as the basic building blocks of meaning. This approach considers the co-occurrence of noun phrases within sentences and builds structure by looking across sentences for those that overlap. Because the sentence is a complete unit of thought (Carley, 1993), mapping noun phrases captures the conceptual units that make up sentences.

Mapping meaning can occur at multiple levels. In a deep parsing approach, meaning can be mapped as localised structures of dependencies. This speaks to a noun phrase as a subject acting upon an object in a single sentence. The same noun phrase may appear in a different sentence as well, but use an entirely different verb phrase to act upon the same object in another. When mapping large discourses, this

may lead to fragmented networks with very little overlap. Therefore, deep parsing is difficult to commensurate into large structures. The shallow parsing approach effectively extracts noun phrases from sentences. By considering the co-occurrences of these noun phrases, it is possible to link sentences through overlapping noun phrases. Therefore the shallow parsing approach is likely to be optimal for generating scalable concept networks across larger corpora of text.

3.5.4 Simple example of a concept network

In this section, I develop a simple example to illustrate how this novel method works. Using a shallow parsing approach, this generates a concept network from a text corpus. To demonstrate how the method creates a concept network from a simple block of text, I consider an example taken from the corpus of tech blog posts about the *tablet*. This particular sample is from a post containing a rumour about the upcoming Apple ³. This *sample tablet rumour* contains two sentences with a total of 61 words,

While a number of rumors have previously pegged the display on Apple's device in the 7-to-10 inch range, this report of a possibly slightly larger display than previously thought nearly matches recent claims of a 10.7" screen.

Speculation that Innolux may provide screens for the new device conflicts, however, with earlier reports that Wintek had already been tabbed as the display provider.

³ Can be accessed at: <http://www.macrumors.com/2009/10/07/foxconn-to-manufacture-tablet-for-q1-2010-launch>

Analysing this with an unsupervised machine learning method, I should be able to derive a meaning structure, however this will not speak to the latent political and situated meaning of the discourse⁴.

The concept network generated by this novel method that corresponds to this *sample tablet rumour* is shown in Figure 3.1. In this figure, there are six edges connecting five nodes in an undirected manner. The node labels are one or more words. The nodes are not fully connected, but rather reflect two broad clusters, that have been coloured in the visualisation. The nodes labelled “screen” and “report/earlier reports” appear as larger and connect the two broader clusters. Each node roughly represents a noun phrase, however this is not an exact correspondence due to their redundancy. I will cover this in further detail below.

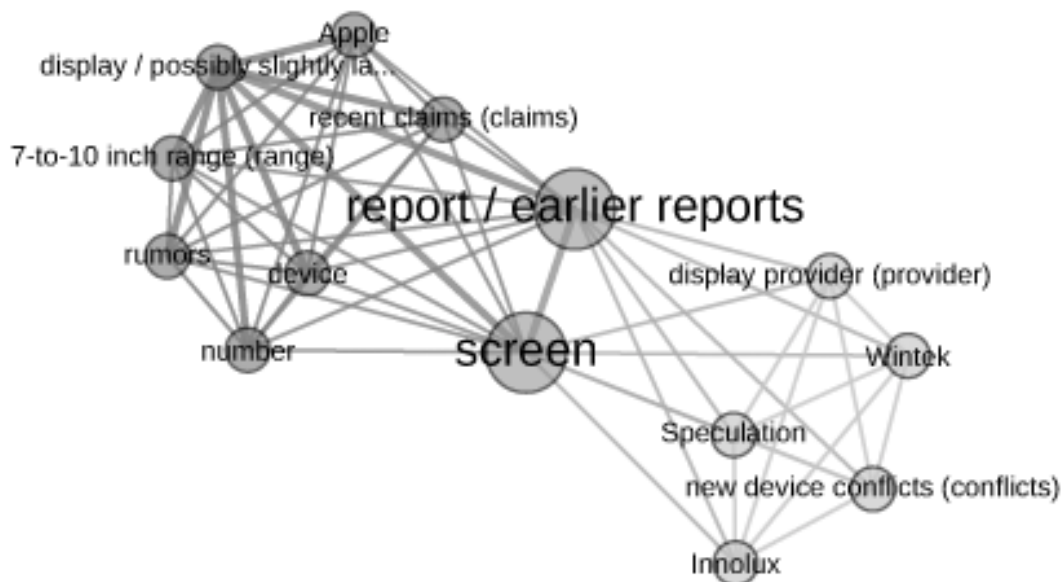


Figure 3.1: Full concept network for *Sample Tablet Rumour*

⁴ This is a noted limitation of content analysis as well, in that analysis is based on explicitly stated concepts in text. However, during the struggles over meaning, there is good reason to think concepts will be made explicit

The method used to derive this network draws upon linguistics and grammars as a means to find structure in text. Before I can describe how this was projected as a network of noun phrases, I first need to discuss how this was processed to extract them. The general steps for finding noun phrases are as follows: 1) extract sentences as separate objects; 2) separate, or tokenise, each word in the sentence; 3) parse and tag each word according to its linguistic Part Of Speech (POS); 3) gather POS sequences into patterns or chunks; 4) find those patterns that are known as noun phrases. This is a simple application of an active area of study known as natural language processing (Bird et al., 2009; Manning & Schuetze, 1999).

Natural Language Processing (NLP) consists of a family of methods that use probabilistic statistical models to infer how language is constructed. These have been developed within the disciplines of Computer Science, Computational Linguistics and Information Retrieval (Manning & Schuetze, 1999). There are many common tasks that NLP can do, but the main goal is often to identify patterns in language to make inferences about its structure in use. As the name implies, it is aimed at using these models to make probabilistic inferences about language as it is used “in the wild.” These models are built into unsupervised machine-learning that processes text along a workflow that hand off to one another. The NLP topics I cover in this section are *tokenization*, *parts-of-speech tagging*, *noun phrase limited decomposability as concepts*, and *grammar parsing for noun phrase extraction*.

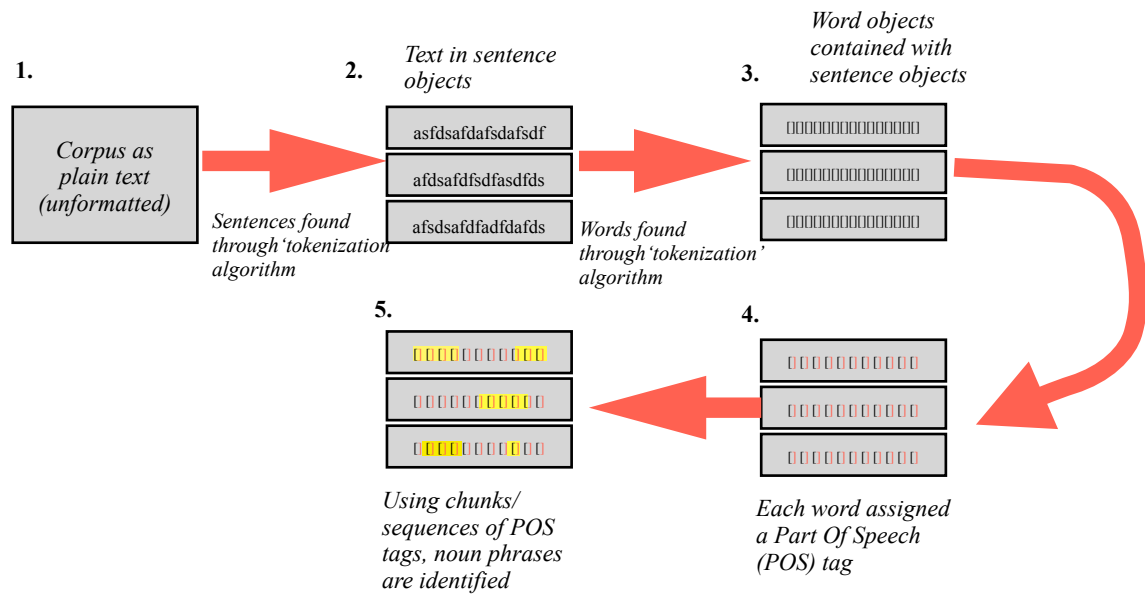


Figure 3.2: Natural Language Processing workflow

The NLP workflow is based around finding patterns within text. In order to identify such patterns, words must be identified and abstracted according to their function within the containing sentence. A common process is summarised in Figure 3.2.

Parts of speech refers to a series of linguistic tags that correspond to linguistic features of English grammar. These include *adjective*, *proper noun singular*, *adverb*, *verb base form* along with many more. Patterns of text can be sequences of parts-of-speech that commonly occur across sentences. In this method, I focus on noun phrases as patterns, or “chunks” that usually indicate the objects in a sentence (Bird et al., 2009). The noun phrase often functions as a subject, object or complement. Noun phrases are more likely than single words to be used to represent complex concepts. The process of searching for such patterns is referred to as “chunking.”

As shown in step 4 of Figure 3.2, identifying parts-of-speech is an important step in the process of chunking. In this method, I adopt the standard Penn Treebank project (see Appendix 2) category system of 36 parts-of-speech tags that are represented by two to four character codes. In this coding system, *JJ* stands for adjective, *NN* stands for a singular noun, and *NNPS* stands for a plural proper noun. This process of tagging parts-of-speech is generally automated by software that has been trained to corpora that have been tagged manually. The software frameworks used in this chapter have been trained according to the CoINN2000 dataset of *Wall Street Journal* articles, towards an 92.5% level of accuracy on POS tagging (Daelemans et al., 1999).

Chunk structures are sets of noun phrases sequences. One example is where multiple nouns appear consecutively. However, grammar is also hierarchical and can be recursive, as evidenced by pronouns. Therefore, noun phrases can be found at multiple levels in a hierarchy.

This grammar hierarchy can be demonstrated when seeing chunk structures visualised as parsed grammar trees. Figures 3.3 and 3.4 show parse trees for the grammar in each sentence from sample *tablet rumour*. These illustrations of parse trees demonstrate how both deep and shallow parsing work. The parse trees show the grammatical structure for each sentences organised in a hierarchical manner. The deep parsing can be read by starting from the root level of the tree and working down. Each major branch will relate to another branch. This partitions the sentence into its constituent parts from the highest level. Depending on what level is considered in the hierarchy, a group of sub phrases can together act as a phrase in

the grammar. To identify noun phrases at the unit level, I use *shallow parsing*. This can be likened to looking bottom of each leaf in the tree, then traverses until the first incidence of a branch of an NP (noun phrase) tag. These represent the smallest possible noun phrase and have been outlined as red boxes in the figures.

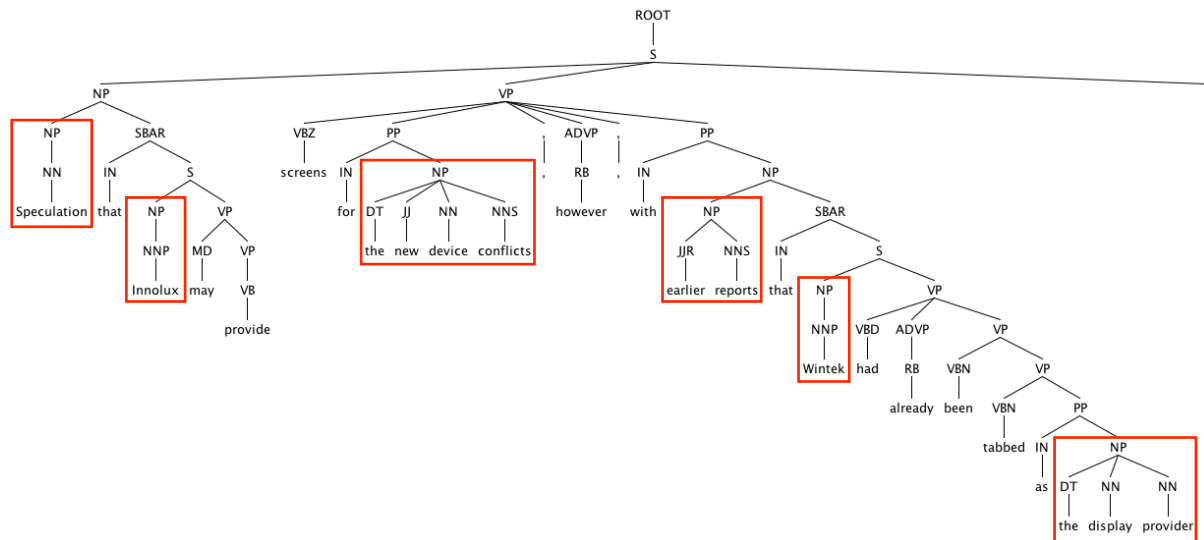


Figure 3.4: Parse tree for second sentence of *sample tablet rumour*

By processing the text through different stages, the method is able to identify linguistic chunking structures that include noun phrases. This is an important point for the method of projecting text as concept networks.

3.6.5 Natural language processing workflow: POS tagging & chunking

Projecting text into concept networks with this method requires the extraction of shallow noun phrases. These are the unit-level groupings that form nodes and their co-occurrence in sentences are the basis for network ties. In this section, I will demonstrate how noun phrases are identified and extracted using linguistically derived Parts of Speech, which are then grouped into chunking patterns. For clarity, I will continue to use the simple example of the *sample tablet rumour*.

The input in the starting point of this process is text that has been organised into sentence form. Most written discourse is likely to already be written in this form, however it is a precondition that the method requires in order to function properly.

Thus, the following text is the input to the first step:

"While a number of rumors have previously pegged the display on Apple's device in the 7-to-10 inch range, this report of a possibly slightly larger display than previously thought nearly matches recent claims of a 10.7" screen.

Speculation that Innolux may provide screens for the new device conflicts, however, with earlier reports that Wintek had already been tabbed as the display provider."

In this analysis, the first step is to separate, or *tokenise* this text into sentence objects. Normally, sentence tokenising is a matter of finding periods, however there are some cases such as *"10.7 screen"* where a period (.) does not signify a sentence full stop. Interpreting the rules and exceptions for sentence separation is a nontrivial task. This work is handled by the NLP frameworks underneath my software and outside the scope of this chapter. Therefore, the first step tokenises the sentences into the following two-item list:

```
Sentences = ["While a number of rumors have previously pegged the display on Apple's device in the 7-to-10 inch range, this report of a possibly slightly larger display than previously thought nearly matches recent claims of a 10.7" screen.",  
             "Speculation that Innolux may provide screens for the new device conflicts, however, with earlier reports that Wintek had already been tabbed as the display provider."]
```

The two sentences of text have been split into sentence objects contained in a list.

By containing each sentence in a list item, the method can operate on each sentence independently and in order. Next, the method tries to find individual words

by applying a word-tokenising algorithm that attempts to find word boundaries. Each sentence object is run through the tokenising algorithm to output the following:

```
Sentences_tokenised =
[['While', 'a', 'number', 'of', 'rumors', 'have', 'previously', 'pegged',
'the', 'display', 'on', 'Apple's', 'device', 'in', 'the', '7-to-10', 'inch',
'range,', 'this', 'report', 'of', 'a', 'possibly', 'slightly', 'larger',
'display', 'than', 'previously', 'thought', 'nearly', 'matches', 'recent',
'claims', 'of', 'a', '10.7"', 'screen', '.'],
[Speculation', 'that', 'Innolux', 'may', 'provide', 'screens', 'for', 'the',
'new', 'device', 'conflicts,', 'however,', 'with', 'earlier', 'reports', 'that',
'Wintek', 'had', 'already', 'been', 'tabbed', 'as', 'the', 'display',
'provider', '.'] ]]
```

Each sentence is contained within brackets and words are separated as individual items. Because the sentences are containers of word objects that have been separated, the method can process each word, but contain it with the context of the sentence. This enables the following step of tagging Parts of Speech. Running that algorithm assigns a POS tag for each word and outputs the following:

```
Sentences_tokenised_tagged = [['While/IN', 'a/DT', 'number/NN', 'of/IN',
'rumors/NNS', 'have/VBP', 'previously/RB', 'pegged/VBN', 'the/DT', 'display/NN',
'on/IN', 'Apple/NNP', 's/POS', 'device/NN', 'in/IN', 'the/DT', '7-to-10/JJ',
'inch/NN', 'range/NN', ',/,', 'this/DT', 'report/NN', 'of/IN', 'a/DT',
'possibly/RB', 'slightly/RB', 'larger/JJR', 'display/NN', 'than/IN',
'previously/RB', 'thought/NN', 'nearly/RB', 'matches/VBZ', 'recent/JJ', 'claims/
NNS', 'of/IN', 'a/DT', '10.7/CD', 'screen/NN', './.'],
[Speculation/NN', 'that/IN', 'Innolux/NNP', 'may/MD', 'provide/VB', 'screens/
NNS', 'for/IN', 'the/DT', 'new/JJ', 'device/NN', 'conflicts/NNS', ',/,',
'however/RB', ',/,', 'with/IN', 'earlier/JJR', 'reports/NNS', 'that/IN',
'Wintek/NNP', 'had/VBD', 'already/RB', 'been/VBN', 'tabbed/VBN', 'as/IN', 'the/
DT', 'display/NN', 'provider/NN', './.'] ]]
```

At this stage in the analysis, the sentence have been separated and each word assigned with a POS tag. From this list, it is also possible to manually identify other parts of speech in the text. For example, going through the list, the verbs that were tagged are:

```
Verbs = [('have', 'VBD'), ('pegged', 'VBN'), ('matches', 'VBZ'),
('provide', 'VB'), ('had', 'VBD'), ('been', 'VBN'), ('tabbed', 'VBN')]
```

At this point, the nouns have also been identified in various forms. However, it is intuitively clear that along with the preposition “*the*”, the nouns “*display*” and “*provider*” are functioning together as a noun phrase. This is why the next step of noun phrase chunking is important. This list of tags enables the chunking process by finding Parts of Speech sequences that correspond to NP patterns. The following is the output from the NP chunking algorithm:

```
NounPhrases = [[('a number'), ('rumors'), ('the display'), ('Apple's'),
('device'), ('the 7-to-10 range'), ('this report'), ('a possibly
slightly larger display'), ('previously thought nearly'), ('recent
claims'), ('a 10.7' screen')], [('speculation'), ('Innolux'),
('screens'), ('the new device conflicts'), ('earlier reports'),
('Wintek'), ('the display provider')]] ]
```

This is the full list of noun phrases that have been extracted from the text. Each one represents a unit level concept. In order to project this as a concept network, the method uses this list and then iterates back over each sentence to see which noun phrases are occurring together. The co-occurrence of noun phrases in sentences is the basis of network tie formation. However there are certain issues around commensurability that I will discuss next.

In this example of the *sample tablet rumour*, the method found two similar noun phrases “*this report*” and “*earlier reports*.” Clearly these noun phrases are referring to a similar thing, however they do not appear exactly similar lexically. This is particularly important when considering how this noun phrase is overlapping the two sentences. Therefore, the method must have a means to take into account a certain level of noun phrase similarity. In order commensurate similar noun phrases, I look at the Dice coefficient (De Smedt & Daelemans, 2012). This simple statistical test

breaks words down into characters. Then it does a comparison based on the number of shared bigrams and gives a measure between 0 and 1. Looking at the two noun phrases, “*this report*” and “*earlier reports*”, we can safely drop the determiner *this* without affecting the meaning. I can then consider the simplified *report* and *earlier reports*. They have ten common bigrams out of a total possible 19. The possible bigrams in each word, including spaces are: *report*: ‘re’, ‘ep’, ‘po’, ‘or’, ‘rt’; *earlier reports*: ‘ea’, ‘ar’, ‘rl’, ‘li’, ‘ie’, ‘er’, ‘r ’, ‘ r’, ‘re’, ‘ep’, ‘po’, ‘or’, ‘rt’, ‘ts’. The similarity measure is calculated by summing up the total number of shared bigrams (5+5) and dividing by the sum of the total number of bigrams (5+14). There are 10 shared bigrams and 19 total bigrams, therefore the similarity score = $10/19$, or 0.53. This measure is important for collapsing similar noun phrases into a node for the concept network.

To continue the example of the *sample tablet rumour*, I now have a list of all noun phrases extracted from each sentence. The method can then look at which noun phrases occur together in each sentence. Figure 3.5 shows how the first sentence from the *sample tablet rumour* forms a simple nine-node network that is fully connected. Because all noun phrases appear in the sentence, they are linked to one another. Similarly, Figure 3.6 shows a projection of the second sentence as a seven-node network.

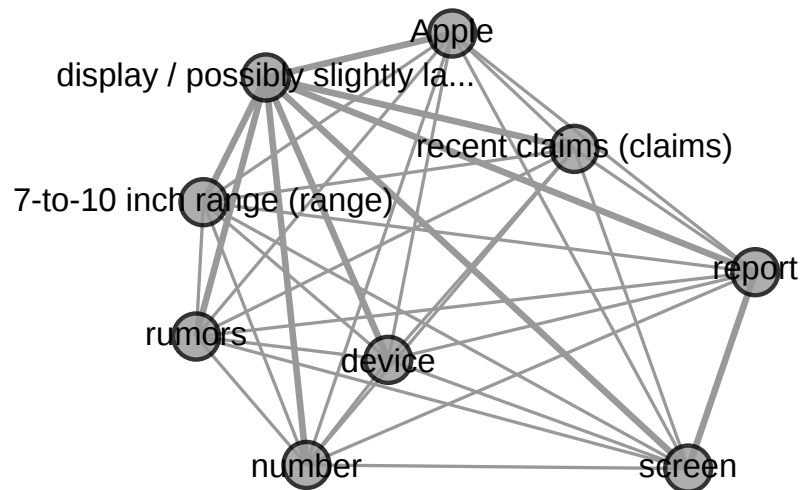


Figure 3.5: Concept network for Blurb Sentence 1

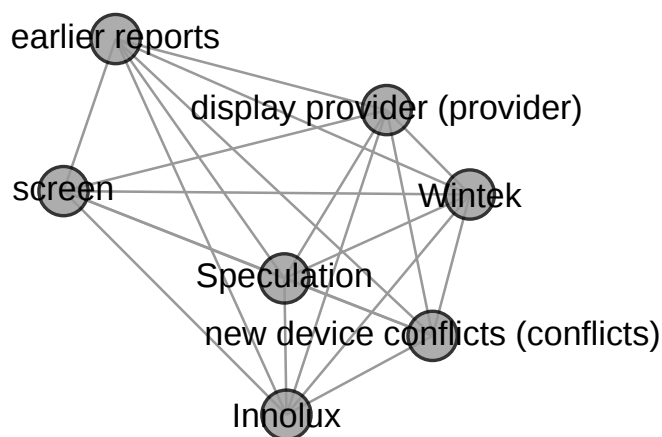


Figure 3.6: Concept network for Blurb Sentence 2

The graphs in Figures 3.5 and 3.6 show each sentence of the *tablet rumour* projected as a simple concept network. These can be linked by finding nodes that overlap both networks. In this case, the nodes labelled “*report*” and “*earlier reports*” both refer to the same head node, *report*. When *report* and *earlier reports* are compared, they have a *Dice similarity score* of 0.53. Interpreting this score as a

means of collapsing noun phrases into nodes is a fuzzy technique that is considered relative to the length of each noun phrase.

The method is able to collapse noun phrases into the same node if their similarity score is above a user-set threshold value. How this threshold is set will affect the type of network projected. In this example, I set the threshold value at 0.5. Because the two noun phrases mentioned above have a similarity score above this threshold value, they are collapsed into the same node. Each noun phrase connection, thus, is aggregated. If there are repeated co-occurrences between the same two noun phrases, this is captured by incrementing the edge weight by one.

Other parameters in the method surround cleaning up the network by dropping nodes that are considered to be noise. This can be particularly important when the research question is looking for robust patterns. One parameter is called *degree pruning*. This is a threshold that drops any nodes that have a degree value below the parameter. For noun phrases on the periphery that rarely appear in the corpus and do not collapse into other nodes, this threshold will drop them. Another related parameter is the *noun phrase frequency*. This is also a parameter to the method that will leave out any noun phrases that appear less than a threshold value. This can also be helpful for tuning the method to focus on repeated noun phrases that make up a nomenclature. It should be noted however, that extreme values of both *degree pruning* and *frequency* will dramatically affect the projection of the concept networks. Therefore, their usage should be sensitised to the research question being considered.

In Figure 3.7, we see the final concept network projection that comes from finding the overlapping nodes between each sentence. Here, the nodes “*screen*” and “*report / earlier report*” have been sized relative to their betweenness centrality (Freeman, 1977). This network measure looks at a particular node connects paths of other nodes. This is akin to the structural hole (Burt, 2004) in social network analysis. Here, the centrality of a node reflects its connectedness in the text. In interpreting this, I draw upon the *centering theory* (Grosz et al., 1995) of linguistics that suggests more central utterances being important for overall meaning of texts. Therefore, the more that a noun phrase appears as a “connector” of other noun phrases, the greater its betweenness centrality and its level of importance the text. Indeed, “*screen*” and “*report / earlier reports*” are the most central nodes and also refer to the meaning of the text.

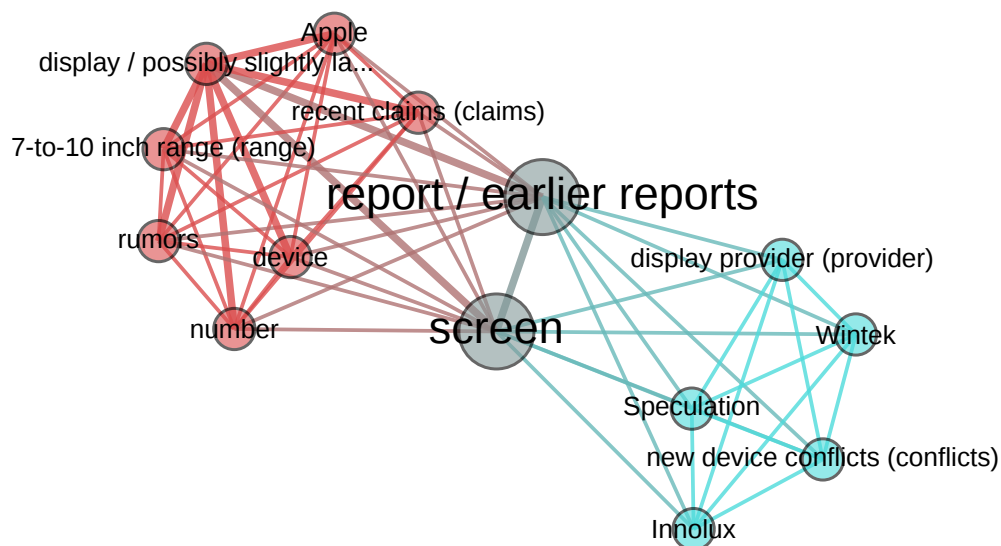


Figure 3.7: Concept network for Sample blurb

In this example illustrated by Figure 3.7, the overlapping noun phrases were collapsed into a single node because the Dice similarity score was 0.53, which was greater than the threshold value set at 0.5. If the threshold is set at 0.51, these noun phrases would not collapse into a single node and the network would appear as two separate clusters reflecting Figures 3.5 and 3.6. To effectively project concept networks from text, it is important to find a meaningful similarity threshold value. Too high a threshold value would mean very little noun phrase collapsing into nodes, only really capturing noun phrases that appeared identical. Singular and plural forms of noun phrases should be collapsed into the same nodes. However, too low a threshold would potentially collapse entirely different noun phrases. As this method is being proposed for use on natural language text, this fuzziness is important. However, there needs to be a sensitivity consideration in finding the right threshold value for similarity. If the threshold similarity parameter is set too low, then the network will be a dense clique. Conversely, if the threshold similarity parameter is set too high, the network will be more sparse with few noun phrases collapsing into nodes. This leads to the proposition,

Proposition 2: *given the right setup parameters, the overlap of noun phrases across sentences projects the structure for a concept network.*

3.5.6 Advanced rules for method in collapsing and labelling nodes from noun phrases

The example in the previous section laid out how to project the *sample tablet rumour* into a concept network. In order to find meaningful overlapping nodes, a fuzzy string comparison was applied to collapse noun phrases into nodes. The noun phrases “*report*” and “*earlier reports*” were collapsed into a single node. For the illustration, the label was listed as “*report / earlier reports.*” However, the method has a more advanced way of inferring the label of a collapsed node. This is done by looking at the head word of each node phrase when calculating the similarity statistic.

Examples of this can be found with nodes “recent claims (claims).” Here, the head word is *claims*. When constructing the concept network, the head word of each noun phrase is compared to the head word of each node. If the similarity is above the threshold, then the noun phrase is collapsed into the node and the label is saved in a list associated with the node. The similarity laid out here is relatively straightforward, however, the heuristic for collapsing nodes and labels has several exceptions around conditions such as proper nouns. These rules handle how labels are derived from collapsing several similar noun phrases. For these advanced rules, more is listed in the appendix (see Appendix 3).

3.5.7 Product ontologies as concept networks

Thus far, I have demonstrated how a concept network can be constructed. Next, I will focus on how product ontologies can be operationalised as concept networks. In order to consider how product ontologies can be operationalised empirically in discourse, I employ a relational view that corresponds to a meaning structure (Mohr, 1998). The cognitive structure of a product market contains several levels including, *product ontology*, *boundary beliefs*, *industry recipe*, and *reputational rankings*. Porac et al., (2002) outline how these levels form a belief hierarchy where lower order beliefs stabilise and provide a platform for those above it. Thus, a belief hierarchy will contain a cognitive network of market actors (Krackhardt, 1987) such as a producer network (White, 1981; Kennedy, 2005). However, at the base of this hierarchy is the product ontology, and this ultimately supports the entire structure. By providing a stable cognitive platform, this claims a space through essential cognitive associations. At this level, cognition is not about a strict definition, but rather a meaning structure from a web of concepts that make up an essential nomenclature. Thus, I propose that an operationalisation of an empirical product ontology can be made by projecting market discourse as concept networks. The same core concepts will likely be linked through overlapping sentences to form a cohesive nomenclature. Like any network structure, a concept network analysis can take advantage of standard network analytic routines (Breiger, 2000; Mohr, 1998).

3.5.8 Product ontology as sets of cognitive classes

The product ontology is a cognitive structure that enables commensurability in a product market. The basis of this idea comes from work on product markets as conceptual systems with attributes embedded in stimulus arrays (Rosa et al., 1999). Porac et al., (2002: 583) extended this work to consider how the product ontology provides the basis for shared mental models by linking,

“usage conditions, product cues, buyer/seller profiles into a nomenclature that distinguishes one product market from another”.

This incorporates the *essential attributes* of a product as an array, but also considers how this cognitive structure forms a nomenclature. This can be thought of as the essential dimensions that enable comparison between products within the same category. For example, to compare one laptop computer to another, I could use the dimensions of *screen size*, *processor speed*, *RAM* and *operating system*. In order for these dimensions to effectively be used to compare products within the category, there first needs to be a stable definition of the *laptop computer* product ontology which determines that those are essential features. A product missing those essential features will not easily fit into the category, and this ambiguity may subject it to an *illegitimacy discount* in value (Zuckerman, 1999). However, the product ontology is also said to include *usage conditions* and *buyer/seller profiles*. Therefore, the nomenclature is not limited to simply an attribute array and could also include other dimensions.

In order to think about a product ontology empirically, I first consider how the attributes form an array around the product definition. To think of this conceptually,

there are a series of concepts that make up the definition. However, the concepts are not all equally as important in a cognitive mapping. At the centre of the definition will be a label (Rosa et al., 1999), and there will be an ordering of concepts that range in importance around it. When thought of as a web of meaning (Geertz, 1973), there is an embeddedness in this structure of concepts. The most central concept will be the label, and other essential concepts will also have central positions. Thinking in this way speaks to relational patterns, where meaning comes from “the systems of relations within which they are contained” (Kirchner & Mohr, 2010:556). Therefore, the meaning of a product is rooted in an understanding of how the attributes work together as a system. This leads to,

Proposition 3:

A stable empirical product ontology will be a nomenclature that corresponds to a concept network held together by broker nodes that link smaller modules to form a structure with the label in the core.

Porac et al., (2002) loosely described the aspects of product ontology as *usage conditions, product cues, and buyer/seller profiles*. To reconcile these components with relational view of a concept web, we can think of individual concepts each being categorised as a component. Combining insights from both Porac et al., (2002) and Rosa et al., (1999), I consider that these cognitive classes can be derived to *Product Attributes, Usecases, and Buyer/Seller characteristics*. Doing this allows the organisation of a concept network by constituent elements and I can conduct analyses based on their dynamics.

3.6 Data analysis

3.6.1 Introduction

I analysed the data in this study by looking at structural dynamics of posting activity on tech blogs. This was divided into two studies corresponding to the derived sub-research questions, *are product publics more likely to be found on the front stage or backstage of online discussion forums*, and *how will a meaning structure stabilise relative to a field configuring event*. In attempting to find evidence of product publics, the goal was to use online discussions to infer mechanisms of collective dynamics (Watts, 2007). By looking at tech blogs discussing the tablet, this observed the proposed phenomena in the natural environment where it was enacted. By studying online interaction data, I was able to capture the transcript of the phenomena as it unfolded along a time line, with all posting activity timestamped. The aim was to establish whether discussions in the front stage or the backstage of tech blogs were more conducive to product public behaviour. This then fed into the second question of where to sample for an emerging meaning structure of a nascent product ontology. The goal of the second question was to find whether a meaning structure emerged in advance of the field configuring event of January 27, 2010.

3.6.2 Analysing and measuring structural dynamics in a user forum

The first analysis looked at a 13-month sample covering a complete snapshot of activity on *macrumors.com*. Taken as a whole, this blog is a discussion forum of interactions in threaded conversations. Each thread is assigned by the site into a thread category. In the data mining process, all posts were organised into threads, which sit within such categories. Posts were gathered and indexed by their threads, along with their authors. I conducted the analysis by looking at structural features of discussion dynamics to infer organising principles.

The literature on networks and information flow assumes that the structure of the network determines exposure to information. Kaltenbrunner et al. (2009) and Gonzalez-Bailon et al., (2010) considered threaded conversations as discussion networks. Both studies found that discussion networks in threads exhibited structural features that corresponded to topics. For example, Gonzalez-Bailon et al., (2010) found that political discussion networks tended to cluster around wider and deep structures, which contrasted with those of discussion networks around religion or Apple. I used this insight to look for structural signatures of communities and publics according to discussion categories laid out by the site. Of particular interest was looking for whether these signatures were more present on the front pages, as determined by the discussion category News and Article Discussion.

This analysis was divided up into three smaller analyses following the methodology of Kaltenbrunner et al. (2009). The first test looked at the homogeneity of activity across thread categories. This could provide evidence to support the proposition that

public and community like activity aggregate to discussion network structures that can occur in different sections on the same site. The second test looked at the overlapping activity of users. Here, the assumption was that thread categories with community like activity would have similar users posting across them. In contrast, a public would likely not have the same users posting across thread categories.

Following the definition of temporary involvement, public-like activity would be for users not to post across thread categories. The third test looked at timing patterns from posts within thread categories. This considered the first condition of a public being ephemeral, as compared to community being sustained for longer periods.

Top level category	2nd level category
Apple Applications	Mac Applications and Mac App Store
Apple Hardware	Apple TV and Home Theater
Apple Hardware	Buying Tips and Advice
Apple Hardware	Desktops
Apple Hardware	Mac Peripherals
Apple Hardware	Notebooks
Apple Hardware	PowerPC Macs
Apple Systems and Services	Mac Basics and Help
Apple Systems and Services	Mac OS X
Apple Systems and Services	Programming
Apple Systems and Services	Windows on the Mac
Archive	Archives of Old Posts
Archive	Wasteland
Mac Community	Community Discussion
Mac Community	Picture Gallery
Mac Community	Site and Forum Feedback
News and Article Discussion	DailyTunes Song Discussion
News and Article Discussion	Mac Blog Discussion
News and Article Discussion	MacBytes.com News Discussion
News and Article Discussion	MacPolls.com Discussion
News and Article Discussion	MacRumors.com News Discussion
Special Interests	Apple Collectors
Special Interests	Digital Audio
Special Interests	Distributed Computing
Special Interests	Visual Media
iPhone, iPod and iPad	iPhone
iPhone, iPod and iPad	iPod

Table 3.3 Thread Categories on macrumors.com.

Table 3.3 shows the top two level of hierarchy of thread categories on *macrumors.com*. These categories are hierarchically organised up to three levels deep and each thread is assigned to at least level two. Several top-level categories have many second level branches. For example, *News and Article Discussion* is a top level category that contains the sub categories of *DailyTunes Song Discussion*, *Mac Blog Discussion*, *MacBytes.com News Discussion*, *MacPolls.com Discussion*, and *MacRumors.com News Discussion*. The subcategory *MacRumors.com News Discussion* includes all the front page articles and their comments.

The analysis compared user and posting dynamics across thread categories to make inferences about publics or communities are constituted on the front pages of the site or back pages of the discussion forum. Following the methodologies of Kaltenbrunner et al. (2009) and Gonzalez-Bailon et al. (2010), I measured three aggregates across thread categories: *proportions of threads and posts*, *user activity overlap*, and *thread timing dynamics*. The thread categories were then compared to find profiles or public like or community like activity.

3.6.2.1 Analysing proportions of activity across thread categories

The first step in this analysis was to look at the proportions of posts, users and threads. The public and community constructs do not suggest greater or lesser proportions, however this first test is a simple descriptive measure to see if the thread categories are equally distributed. If this were true, this would speak to homogeneity on the site. Otherwise, this would be further evidence to support a

different set of organising principles present within the site. I conducted this analysis by looking at descriptive statistics across thread categories, and compared normalised proportions.

3.6.2.2 User overlap

The second step for the analysis was to look at user participation across thread categories. The goal of this measure was to find and compare profiles of user participation. Each thread category is made up a set of threads. By looking at individuals post for each thread, I tracked how often a user was participating. When aggregated to the level of a thread category, this provided a usage participation profile. After constructing each thread category profile, I then compared across them to find how similar or different they were.

To build a user participation profile for a thread category, I constructed a user participation vector (Kaltenbrunner et al., 2009). In this vector, each dimension represented a user and recorded how many times they participated in threads that resided in this category. Because there are 8 thread categories and 78,785 unique users, this meant constructing 8 participation vectors, each with 78,785 dimensions.

Comparing the vectors was performed by normalising, and then contrasting them using cosine similarity calculations. This is a common method used in the information retrieval field (Manning et al, 2008) and implemented in the software framework *Natural Language Toolkit* (Bird et al., 2009). The outcome of this is a matrix with 8 rows and 8 columns. Each cell represents a number comparing two vectors. By constructing a vector for each user, then cycling through thread

categories for involvement. Each vector is normalised using TF-IDF and a distance measure is constructed comparing cosine similarity. Darker colours indicate more similar users getting involved and lighter shades are less similar.

3.6.2.3 Timing dynamics

The third step in analysing thread categories was looking at timing dynamics of threads. In a thread, the difference between the first post and the second post is a time delta. The difference between the first and third post represents another time delta. Each thread will have a distribution of time deltas that indicate a shape of conversation dynamics. Threads where the majority of posts come in within a short time will have a curve with a strong positive skew. This is known statistically as being skewed the right. For threads that remain dormant for some time and then pick up quickly before dying off, their distributions will show a strong negative skew, or having a left skew. I analysed the time deltas each post in per thread. I then aggregated these at the level of thread categories. I then used box-whisker plots to demonstrate distributions and compared the skewness and kurtosis values of each.

3.6.2.3 Measuring public and community activity

Community has intuitive analytical appeal as a structural form. The most active members in the core are also the most visible, and the regularity of their presence goes a long way towards sustaining concepts of norms, precedent, and identity. Importing the sociological version of community from an offline social world has not been without problems (Preece & Maloney-Krichmar, 2005), but it has indeed become the primary framework for much research into online activity. One alternative

is to consider the public as a means to analyse dynamic discursive activity. As a classic construct in sociology, a *public* has been given many treatments (Dewey, 1927; Blumer, 1948; Shibutani, 1966; Goffman, 1974; White, 1995; Fine, 2010). The central tie amongst them is that of a temporary arrangement of social actors enacting a discursive activity.

When considering the question of organising principles of activity distribution, the frame of community grants primacy to social relations. That is, in thinking about posting activity, we should think that social relations take on a form of causality. In an online forum, users who know one another should have a tendency to post in the same threads. Faraj & Johnson (2010) made this point about sustaining online communities through network exchange patterns.

Under the question of different organising principles, we can explore what underlies users' propensities to post in threads. From Figure 3.8, I consider a spectrum whereby a user will respond to a thread either because they have an existing social relation with other posters, or simply an interest in the topic being discussed.

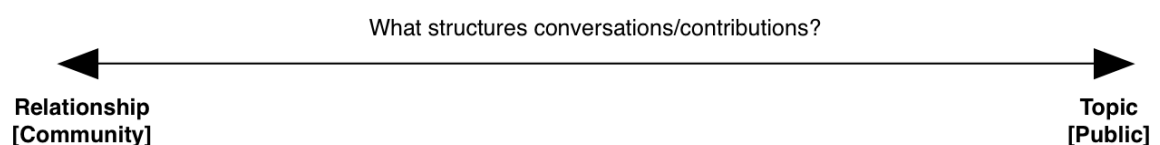


Figure 3.8: Spectrum of mechanisms explaining Users propensity to post

The logic of community suggests that users will post together in similar patterns. Due to a social relations, this similar posting could occur across different topics and

threads. Furthermore, they likely will position their posts and responses close to one another. Conversely, interest in a topic should not necessarily have much to do with posting close together. If users have a tendency to post in related threads but with no significant position distance, we can posit that the attraction is to the topic and not to one another. The logic of a public would be coming together to discuss an issue with no discernible pattern of closeness with related users.

Measuring closeness of posting within a thread is a methodological challenge due to the need for a critical value of closeness and an associated sliding sampling window. However, a simple operationalisation of this may suffice for the research question. I propose that by considering which threads users tend to post in and how often, this will provide sufficient evidence for tracing activity signatures of publics and communities.

The three primary characteristics of publics are *ephemeral interaction*, *temporary suspension of identities*, and *topical focus*. Looking at these on *macrumors.com*, I considered three types of analyses to contrast activity across thread categories. First, I looked at the distribution of post and thread proportions to test for homogenous activity. Then I looked at user overlap and timing dynamics to compare patterns across thread categories. Because thread categories correspond to front and back stages, I could then determine where public and community like activity were taking place.

3.6.3 Analysing and measuring concept networks

Thus far in the chapter, I have outlined how I found evidence of public-like activity on the front page of a major tech blog, *macrumors.com*. This then provided the basis for considering the broader sample of front pages from the three other tech blogs in the dataset. I then further filtered this sample based on posts that included the keyword “*tablet*”, and used them to construct a text corpus. From this corpus, I organised posts into sets according to the quarterly date period in which they were published. I then needed to project concept networks to analyse the emergence of meaning structures. This led to the development of a novel method of generating concept networks from text. I will describe here how such a method can be used for analysis.

Earlier in this chapter, I illustrated a novel method for constructing a concept network from a simple text corpus. In this section, I extend these ideas towards a consideration of how they can be measured and analysed. The advantage of working with relational data as networks is that one can follow standard network analytical routines (Breiger, 2000; Mohr, 1998). Networks are not at all limited to social ties (Monge & Contractor, 2006; Watts, 2004) and have been observed across biology, macroeconomics and other settings. However, the social network analysis literature has matured and contributed significantly to sociological and organisational literatures. Therefore, I attempt to extend ideas from social network analysis to the analysis of concept networks.

Analysis in the social networks literature is broadly divided into structural configuration issues such as embeddedness, and the centrality of social actors

(Kilduff et al., 2006). These are interesting frames for analysing concept networks as well. Work on meaning structures speaks to types of discourse being embedded in social contexts (Mohr, 1998; Kennedy, 2008). The centering theory from linguistics (Grosz et al., 1995) says that central concepts are thematically important as connecting nodes. Following these, I consider here three forms of analysis for concept networks. The first analytical tool in a concept network is betweenness centrality (Freeman, 1977) of nodes as bridges. This is a core insight of CRA (Corman et al., 2002) showing individual nouns as keywords as being central to meaning. The second analytical measure is the degree of nodes. Edges between nodes indicates when they co-occur in sentences. The final analytical tool I will consider is the network measure of modularity. This speaks to the level of connectedness amongst clusters in a network.

Doerful & Connaughton (2009: 206) proposed that a semantic network analysis could provide an “assessment of the corpus’ dominant themes using network analysis techniques.” This was operationalised in CRA and the method described here through betweenness centrality (Freeman, 1977). This measure is mathematically derived by looking at shortest path lengths, or the number of geodesics passing through a node (Monge & Contractor, 2006). It can meaningfully be described as how much a node is a *broker*. For concept networks, a broker concept can be interpreted as one that connects other concepts and holds a structure together. This can be intuitively understood when considering how any piece of text is separated into sentences and paragraphs. In order for flow among these structures, a connector between them must exist to hold the structure together.

Therefore, high betweenness centrality can indicate thematic importance of a concept to a text.

The degree of a node measures how many other nodes are connected to it. This can be understood at the level of a node and that of the entire network structure. As a network measure, this enables us to consider degree distributions as evidence of hubs and preferential attachment (Watts, 2004). The degree also can help to discriminate important nodes from less important ones. Considering degree values, a large network can be filtered to only show nodes with degrees above a threshold. This would hide less connected nodes which may appear as noise, and highlight ones that are more densely connected in the network.

3.6.3.1 Modularity as a measure of discursive cohesiveness

The third analytical tool I discuss in this chapter is the network measure of modularity. This is derived from a class of algorithms known as community detection and gives a score between 0 and 1 (Leicht & Newman, 2008). It indicates the degree to which networks have salient groups, or communities, where a community is defined as “a subset of a larger population where internal ties are more prevalent than ties to other subsets” (Shwed & Bearman, 2010: 822). I use this measure on concept networks to find a proxy for consensus in the discourse. A central tenet to the product ontology is that there will be some level of consensus as agreement amongst social actors in a field. This has parallels to the work in sociology on scientific consensus. Since Merton (1979) and Kuhn (1970), we have developed an

understanding of scientific fields converging upon facts, but most studies were domain specific. The quantitative measure of modularity has potential validity in both the product ontology and other cases.

Recently Shwed & Bearman (2010) found a quantitative measure for consensus by using citation networks. They applied modularity to empirically demonstrate when consensus was occurring around topics such as smoking and cancer and the suspected carcinogenicity of cellular phones. I take this insight forward to the study of product ontologies to find when they stabilise and how. To do this, I needed to develop a method for projecting multiword noun phrases with fuzzy-set similarities. In this chapter, I described the details of this method. Below, I show how it can be used to analyse longitudinal data around product ontology emergence.

If there are several "islands" (or salient, disconnected communities) with relatively few broker connections, then that corresponds to a high modularity score. As more broker connections come in, the modularity level goes down. Mathematically, this procedure corresponds to randomly rewiring connections on adjacent nodes and seeing what outcome this has on the network shape. In Figure 3.9, we see the top two cases both have modularity as zero. Rewiring these networks would produce the same shape. On the bottom half of the figure, we see three cases of networks with modularity values reported as the variable Q . On the left side there are two components disconnected from one another. As more broker connections are made between them, the modularity score declines.

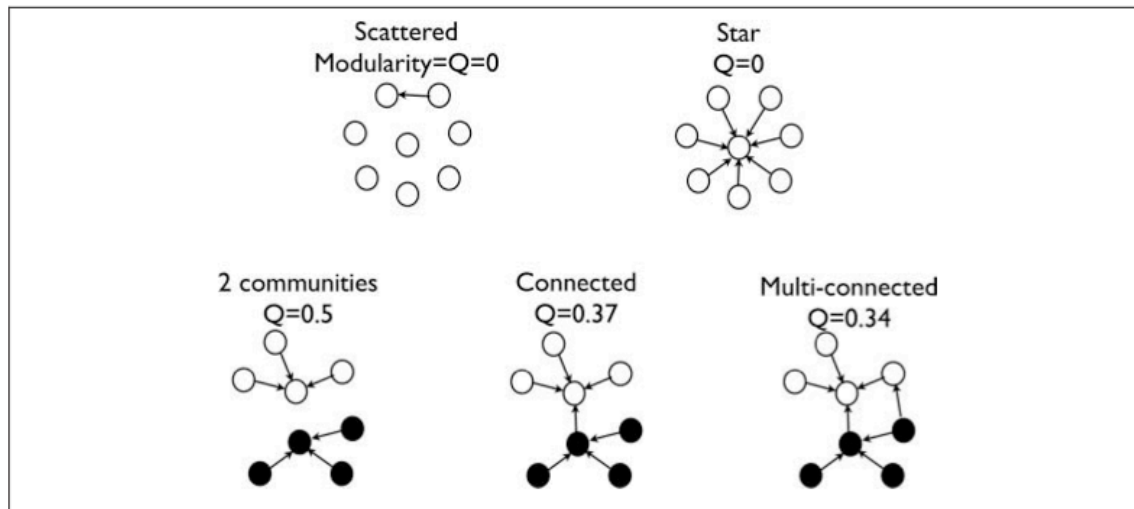


Figure 3.9: Modularity explained (source: Shwed & Bearman, 2010)

One example of how modularity has been used analytically is Shwed & Bearman's (2010) paper on the formation of consensus in science. Sociologists of science have been dealing with the issue of scientific consensus for several decades (Collins, 2004; Frickel and Moore, 2006; Fujimura, 1997; Gieryn, 1996) since Merton's (1979) influential work on how the institution of science functioned. This work covered norms, values and rewards and presented scientific knowledge as a rational, neutral domain divorced from politics or social influence. In this, Merton was more concerned with the functioning of the institution than of the content produced (van House, 2003). Consensus was deemed a natural rational outcome. In this depiction, politics, funding incentives, and culture were ignored. However, Merton may have had good reason to approach the scientific institution as an object of study by maintaining some distance and by not privileging an epistemological position through domain expertise. Overturning Merton's functionalist model of scientific consensus became a major focus for STS (Science and Technology Studies) scholars, who investigated the many mechanisms behind it (Latour 1987; Shapin 1996; Star and

Griesemer 1989). Scientific consensus was viewed as deliberation followed by closure (Pinch & Bijker, 1984). Studies on this were primarily qualitative and limited in generalizability.

Shwed & Bearman (2010) introduced a method for capturing the temporality of consensus formation. They developed modularity as a quantitative analysis of citation networks in scientific publications. By projecting citation networks for all papers in a given period using a series of keywords, they were able to map out a structure of centrally located papers. It changed over time as references to key papers clustered and became ritual cites. The main insight behind the method is that the structural modularity property of a citation network can have explanatory power for scientific consensus.

Extending Shwed & Bearman (2010), I attempt to use modularity as a measure of cohesiveness and consensus for concept networks. Once the network has been constructed, it is passed through Leicht & Newman's (2008) Community Detection algorithm using the iGraph software package implementation (Csardi & Nepusz, 2006). This algorithm will take the concept network as input and calculate a modularity score between 0 and 1. This measure tracks how connected the concepts are in a network, with a score of 1 being maximal modularity. If the meaning structure in a corpus is not cohesive, it may have several well-connected clusters of concepts that do not connect with one another. Further, an incoherent text may contain small clusters that are not well connected to the broader graph. For values close to, but above zero, a low modularity will correspond to a well-connected graph, with many broker concepts. A corpus of texts that correspond to the same topic is proposed

here to have a low modularity score. Thus, the higher “connectedness” of a corpus corresponds to a relatively tighter level of consensus. In line with corpus linguistics as a method most effectively used when comparing corpora (Pollach, 2012), this method is proposed for either tracking the same corpora across longitudinal slices or comparing theoretically interesting cases. The intuition follows Carley’s (1997) insights that consensus and indeed knowledge in a cognitive network is a well-connected web of concepts. This is similar to the Geertz (1973) notion of culture and meaning being embedded in a web, which leads an empirical proposition:

Proposition 4: *For a corpus made up of articles that are qualitatively more topically similar, the modularity scores are lower.*

This proposition can be demonstrated rather simply when considering the following: a noun phrase network captures how “connected” the concepts of an article are. If there are several “islands” (or salient, disconnected communities in network-science) with relatively few broker connections, then that corresponds to a high modularity. As more broker connections come in, the modularity level goes down. Intuitively, a series of articles with similar concepts being discussed in a similar way will have more broker connections, and thus lower modularity. This is not an all-encompassing and flawless measure of consensus, but I believe that it can be used in a similar way as Shwed & Bearman (2010) to sense coherence in a corpus over time. Further, I propose this can be used to map and track product ontologies evolving over time.

3.6.3.2 Analysis of cognitive class centrality in longitudinal study

I consider the aspects of a product ontology as a typology, which enables me to aggregate concepts into broader buckets to trace the process of stabilisation. By conducting this analysis for each quarter in the sample, the method is able to derive centrality scores for each cognitive class in each quarter. To find the relative measures for each cognitive class in a quarter, I divide the “centrality sum” score by the sum of this for all classes. The normalised value can then give relative proportions for each class in each quarter. I can then compare the proportions of quarters over time as a longitudinal analysis.

Such an analysis speaks to the cognitive embeddedness over time. Not only can I consider the web of meaning from the method overlapping co-occurrences noun phrases in a concept network, but this operationalisation should also allow me to trace how the elements of such a structure are assembled over time. I propose this can be done by considering the elements of a product ontology as cognitive classes. Recall that each of these elements are “Product attributes”, “Use cases” and “Buyer/Seller characteristics.” To find how these elements are being assembled in the cognitive structure over time, I code each node into a class, then keep track of its centrality measure. By summing up centrality measures for all nodes within a class, I am able to compare relative centrality across classes. I compare these relative measures across network snapshots over time.

The method will help to track embeddedness of cognitive classes through following relative centralities across quarterly network projections. I code for the base three product ontology elements, however I also consider the additional codes for: *labels of products, labels of categories, time reference and information sources*. Because I am considering when an ontology stabilises, labels may emerge as central cognitive referents which blackbox meaning (Latour, 1987; Grodal, 2007; Shwed & Bearman, 2010). They can also serve as analogies as making sense of new categories is done through a consideration of established ones (Bingham & Kahl, 2012). Time reference and information source are coded to find evidence of a pending product release and rumour activity. Because the corpus runs across a range before and after major products were announced, references to time and the source may also indicate the level of speculation, which may be a mechanism for stabilisation.

For this analysis, reported in Chapter 5, I projected concept networks for all 20 quarters in the tablet sample corpus from 2006-Q1 to 2010-Q4. The data from 2006-Q3 was too small and was dropped from the analysis. For each concept network, I derived a list of the top 30 nodes arranged in descending order by betweenness centrality measures. Each of these were then coded and centrality measures gathered. Table 3.4 shows a sample analysis of 2009-Q3. In this table each node arranged into respective coded groups. The sums of the centrality scores for each coded group are also shown. Because each coded group represents a cognitive class, this gives a measure for the relative centrality of each.

Q3 2009	Node label	Centrality	Frequency	Examples	Cognitive Code
	store	0.007	23	[14 keys] store(x6), e-book store	UC
	month	0.017	41	[21 keys] month(x10), six month	TM
	year	0.012	41	[14 keys] year(x18), early next ye	TM
	time	0.009	35	[16 keys] time(x14), first time(x5)	TM
	event	0.007	30	[12 keys] event(x18), Apple ever	TM
	screen	0.022	39	[18 keys] screen(x15), 6-inch sci	PA
	version	0.021	23	[16 keys] version(x7), 160GB 50	PA
	model*	0.018	26	[18 keys] models(x4), model(x4)	PA
	processor*	0.01	21	[17 keys] 2 Duo processor(x2), F	PA
	camera	0.009	19	[7 keys] camera(x9), digital cam	PA
	Bluetooth	0.007	8	[2 keys] Bluetooth(x7), 3D Maps	PA
	iPod touch (touch)	0.012	32	[22 keys] iPod touch(x10), Jourr	LP
	Window	0.011	24	[5 keys] Windows(x20), Pen Win	LP
	iPhone	0.009	28	[7 keys] iPhone(x22), large iPhor	LP
	tablet / Apple Tablet	0.22	153	[74 keys] tablet(x52), Apple Tabl	LC
	device / new device	0.191	136	[49 keys] devices(x69), new dev	LC
	tablet computer (comp)	0.016	34	[10 keys] tablet computer(x16),	LC
	report	0.058	66	[33 keys] report(x26), Gizmodo	IS
	column	0.018	5	[1 keys] column(x5)	IS
	source	0.014	46	[15 keys] source(x27), industry s	IS
	article	0.008	4	[1 keys] articles(x4)	IS
	Apple	0.214	176	[8 keys] Apple(x169), whether A	BS
	company	0.08	37	[7 keys] company(x29), Cupertino	BS
	technology*	0.026	11	[9 keys] consumer technology(x	BS
	industry	0.022	7	[3 keys] industry(x5), industries.	BS
	price	0.009	17	[9 keys] Price(x7), Prices(x3)	BS
	Microsoft	0.009	24	[3 keys] Microsoft(x22), system	BS
	Archos (Archo)	0.005	37	[6 keys] Archos(x31), Android-b	BS
	tech / passion tech	0.049	4	[2 keys] tech(x2), passion tech(x ?	
	product	0.007	27	[11 keys] product(x17), launch p ?	
	project	0.006	22	[12 keys] project(x10), tablet prc ?	

Cognitive codes name	Code	Centrality Sum
Product Attributes	PA	0.087
Usecase	UC	0.007
Buyer/Seller characteristics	BS	0.365
Label of Product	LP	0.032
Label of Category	LC	0.427
Time reference	TM	0.045
Information Source	IS	0.098
Uncodable	?	0.062

Table 3.4 - Sample of Cognitive Classes coding - 2009 Q3

3.7 Applications of novel method

3.7.1: Introduction

In this section, I report on an empirical analysis demonstrating the method of concept networks. For this, I used posts selected from the *tablet* speculation corpus to show the propositions that have been developed around concept networks. This analysis was conducted by sampling 4 sets of blog posts based on date published and qualitatively determined topical similarity. The goal is to show how modularity changes according to different sampling conditions.

3.7.2: Demonstrating propositions in an analysis

According to proposition 2, the parameter of threshold similarity will shape the structure of the concept networks. The research question here is: *how does the modularity measure shift when the corpus changes to include topical diversity?* The power of any analysis with corpus linguistics is based on comparing corpora (Pollach, 2012). It is sensible to compare corpora derived using the same analytical parameters. I used *similarity threshold = 0.75* to capture a balance between similar noun phrases and an adequate structure. Setting this parameter too low would have collapsed many noun phrases into few nodes at the expense of structure. These would connect as a near-clique. Setting it too high would have collapsed noun phrases into many nodes and resulted in too fragmented a structure. Given the small size of the sample, I kept *frequency* and *degree Pruning* each at 0. This then included all noun phrases.

Proposition 3 was written as being specific to the product ontology, however, this speaks to the broader idea of explicit meaning and consensus in a corpus. The product ontology is a common nomenclature in discourse. A well-formed product ontology should result in a network of well connected modules. However, this is not simply one dense module. Rather, it should have a medium sized modularity score that comes from a series of well-connected dense sub-modules. That is, the elements of a product ontology make up a set of cognitive classes that include, *attributes usecases*, and *buyer/seller profiles*. Each of these classes is likely a sub-module, and loosely connected in a concept network. Therefore, it may also be analytically interesting to consider the list of broker nodes that have high betweenness centrality to find what the core of the product ontology structure is.

According to proposition 4, modularity in a concept network can be a function of topical diversity in a corpus. This suggests that one well-defined topic will correspond to a dense network module. When two well-defined and separate topics are combined in a corpus and projected as a concept network, the resultant network should have two dense modules, and a higher modularity score. Given that modularity is a measure relative to the entire network considered, then the additive effect of combining two networks with low modularity will result in a network with higher modularity. However, if those two topics are similar, a combined network will contain several overlapping nodes. This could result in a lower modularity score for the combined network.

3.7.3: Sample analysis of concept networks from corpus

For this sample analysis, the goal is to show how modularity in a concept network is related to topical diversity of a corpus. To show this, I created four small corpora based on certain conditions, and these were each projected into base concept networks. In constructing the corpora, I manually sampled articles based on topical similarity and date range. Not having a corpus similarity metric, the assumption here is that qualitatively similar articles within the same quarter should contain significant overlap and a low modularity score. I combined each set according to theoretically interesting conditions and recorded the change in modularity score. Each base corpus was sampled and generated similar network profiles, despite the variance in dates published and topics. All networks were projected with the same parameters.

	Network1	Network2	Network3	Network4
Network Type	Single topic, single period <i>"Apple tablet Q409"</i>	Single topic, single period <i>"Dell tablet PC Q407"</i>	Single topic, single period <i>"Apple tablet Q309"</i>	Single topic, single period <i>"Android tablet Q309"</i>
Nodes	99	115	148	136
Edges	481	929	905	1356
Modularity	0.523	0.423	0.440	0.430
Scaled modularity	0.194	0.205	0.203	0.201
Average Degree	9.7	16.16	12.23	19.94
Noun Phrases extracted	237	200	328	210
Total Words in Corpus	938	692	1184	759

Table 3.5 Descriptive statistics of base concept networks

In Table 3.5, descriptive statistics are shown for the base networks. These are based on finding three different topics and three different time periods, with some overlap. Each corpus was constructed from 4 articles and according to a sampling condition. The first condition was looking for articles in Q42009 that contained the theme of *Apple tablet rumours*. Next, I looked for a sample of articles that were around a separate topic in a separate quarter. I sampled articles in Q42007 that had the theme of the *Dell Tablet PC*. The next sampling condition looked for articles similar to the first sample network, but in a different quarter, *Apple tablet Q32009*. Finally, the last sampling condition was to find a similar time range, but a different topic. I sampled articles from *Q32009* that had the theme *Android tablet*. Each of these sample networks had similarly low-scaled modularity scores, indicating well formed topics. Figures 3.10, 3.11, 3.12, and 3.13 each show sample articles from the corpora.

Article: "Apple Asking App Developers to Prepare for Large Screen Demos Next Month?"
Published: Q42009, Dec 23, 2009
Source: Macrumors.com (<http://www.macrumors.com/2009/12/23/apple-asking-app-developers-to-prepare-for-large-screen-demos-next-month/>)

Content:

Silicon Alley Insider reports that Apple has approached certain app developers to ask them to prepare for a demo scheduled for next month. Notably, Apple has requested that the developers prepare their applications to run in "full screen" mode rather than limiting display to the iPhone and iPod touch's 320 x 480 pixels. Apple is preparing to show off a new, larger mobile device with a higher resolution display in January, probably a version of the Apple tablet we've been hearing about for months, according to a plugged-in source in the mobile industry. Apple has been telling some app makers to prepare apps for a demo next month, according to this source. "They've told select developers that as long as they build their apps to support full screen resolution, rather than a fixed 320x480, their apps should run just fine," our source says. The source notes that only the demo will occur in January, with a device launch scheduled for sometime in the future to allow other developers to modify their applications to support the new device and to allow Apple to generate excitement and anticipation ahead of the launch as it did with the original iPhone. Apple tablet rumors have been flying fast and furious lately, with rumors of a February production ramp for a March or April release leading to reports of a possible January announcement. Exact specs of the tablet remain unknown, with long-standing rumors of a 10" display being joined in recent weeks with increasing claims of a 7" display, possibly indicating plans for multiple models of the new device.

Figure 3.10: Sample article on apple tablet rumour, Q42009

Article: "Dell Latitude XT tablet is official, sexy"
Published: Q42007, Dec 11th, 2007
Source: Engadget (<http://www.engadget.com/2007/12/11/dell-latitude-xt-tablet-is-official-sexy/>)

It's absolutely no secret that Dell's been working on a tablet PC -- we've been hearing about the damn thing since as far back as mid 2004 (no joke!) and the company did promise it in 2007, which it's almost not. What took them so long to get it out the door we've no idea, but it's hard to deny the appeal of this little ThinkPad-esque bugger. Here are those specs which you've absolutely never seen before: 12.1-inch WXGA (1280 x 800) in CCFL ("outdoor display," up to 400cd/m2) and LED backlight options Capacitive touchscreen(!) with palm detection (this, if anything, could enable that supposed multi-touch), and active digitizer. Intel Core 2 Solo or Duo ULV at 1.06 and 1.2GHz, respectively ATI Radeon X1250 UMA Graphics Up to 3GB memory, runs Vista Ultimate or Business 40GB or 80GB 4200 RPM drive, 120GB 5400 RPM drive, or 32GB or 64GB SSD HSDPA or EVDO rev A., 802.11a/b/g/n options, Bluetooth 2.0, and gigabit Ethernet connectivity ExpressCard54 and SD slots, 1394, VGA, (3) USB 2.0 Scroll wheel, biometric reader, TPM 1.2, super low profile power adapter Up to 5 hours battery life with standard 43WH pack, up to 9 hours with extended 45WH pack; Energy Star 4.0 compliant 3.57 pounds (presumably without battery), 1-inch thick Starts at \$2500 Okay, we'll hand it to these guys, alongside the M1330 the Latitude XT is the sexiest machine Dell's ever produced -- and the addition of a capacitive touchscreen totally blew our minds. But \$2500 for the most stripped down option? Dell just priced itself right out of the game, if you ask us. Hands-on coming right up.

Dell Latitude XT tablet is official, sexy

Figure 3.11: Sample article on Dell tablet PC rumour, Q42007

Article: "What Will be the Killer Feature of an Apple Tablet?"

Published: Q32009, July 28, 2009

Source: Macrumors.com (<http://www.macrumors.com/2009/07/28/what-will-be-the-killer-feature-of-an-apple-tablet/>)

With the growing evidence that Apple will be releasing some sort of Apple Tablet in the next 3-6 months, analysts and writers have started speculating about the potential success or failure of such a device. Both Gizmodo and PC World have taken the time to detail why they think that such a product will be a flop. John Gruber, however, points out that these people are incorrectly assuming that such a device would mimic present day tablets and offer no compelling new features. This is not the first time that Apple has been close to releasing a tablet device. Back in 2003, the evidence for a Mac Tablet had reached similarly high levels, but for whatever reason, Apple ultimately decided not to release that device. We've always felt that the reason has been more a marketing decision than a technical one. Until Apple comes up with that "killer feature" for a tablet, we don't believe that Apple would commit to the market. Apple has always said that they won't pursue a product line unless they feel they can offer something new and compelling. While these promises could simply be dismissed as marketing propaganda, it seems clear that Apple's entries into the MP3 and mobile phone markets have met those standards. Arguably, the iPhone's killer feature on launch was its excellent mobile web browser. This feature was poached from another internal Apple tablet project called "Safari Pad". Steve Jobs is said to have recognized its value and morphed it into what became the iPhone. So the question remains, what added value has Apple decided it can provide in a tablet device that its competitors have been unable to offer? Are interactive album booklets alone compelling enough to launch this new device? Or have they finally decided to deploy more advanced multi-touch on a larger screen?

Figure 3.12: Sample article on apple tablet rumour, Q32009

Article: "Archos phone tablet teased: Android, 4.3-inch touchscreen, 1GHz ARM processor"

Published: Q32009, Sep 15th, 2009

Source: Engadget (<http://www.engadget.com/2009/09/15/archos-phone-tablet-teased-android-4-3-inch-touchscreen-1ghz/>)

In an all-too brief teaser at a press conference across the pond, Archos decided to give a few more details and a teaser pic of the Android phone it announced back in February. According to the slide, it's got a 4.3-inch touchscreen with 854 x 480 resolution, a 1GHz ARM processor, likely comparable to Archos 5, but pretty powerful considering what all other Android phones have so far arrived with, and a 10mm-thick shell. Size and specs-wise, it matches up quite handily to the Snapdragon- and Windows Mobile-powered Toshiba TG01, although that one's already taking up retail shelf space, and it looks like we'll be waiting for some of those infamous Archos leaks if we want any more information the phone tablet in the near future.

Figure 3.13: Sample article on Android tablet rumour, Q32009

3.7.4: Sample results

Table 3.6 shows the descriptive statistics of the base networks and the derivative combined corpus networks. Corpora were combined using the following conditions: *mixed topic across mixed periods*, *single topic across mixed periods*, and *mixed topics within the same period*. I derived a scaled modularity measure, by dividing the modularity score by the log of the network size. This normalises the analysis and allows the networks to be compared. Networks 1 to 4 in the table are derived from the base corpora and Networks 5 to 7 come from the combined corpora.

	Network1	Network2	Network3	Network4	Network5	Network6	Network7
Network Type	Single topic, single period "Apple tablet Q409"	Single topic, single period "Dell tablet PC Q407"	Single topic, single period "Apple tablet Q309"	Single topic, single period "Android tablet Q309"	Mixed topic, mixed period "Apple tablet Q409", "Dell tablet PC Q407"	Single topic, mixed period "Apple tablet Q309", "Apple tablet Q409"	Mixed topic, single period "Apple tablet Q309", "Android tablet Q309"
Nodes	99	115	148	136	201	216	268
Edges	481	929	905	1356	1390	1341	2247
Modularity	0.523	0.423	0.440	0.430	0.535	0.387	0.518
Scaled modularity	0.194	0.205	0.203	0.201	0.232	0.166	0.213
Average Degree	9.7	16.16	12.23	19.94	13.83	12.42	16.77
Noun Phrases extracted	237	200	328	210	438	565	537
Total Words in Corpus	938	692	1184	759	1630	2122	1942

Table 3.6 Descriptive statistics of sample and combined concept networks

The results of this exercise can be summed in the 2x2 matrix found in Table 3.7.

Each base network has a low-scaled modularity score and corresponds to the cell in the top left. This is the result of each adhering to a well-defined topic. The reported scaled modularity scores for networks 1-4 are *0.194*, *0.205*, *0.203*, *0.201*.

	Single topic	Mixed topic
Single Time period	<i>Low</i>	<i>Medium</i>
Mixed Time period	<i>Medium-Low</i>	<i>High</i>

Table 3.7 Effect of modularity from variance of sampling conditions

When I varied the period and mixed separate topics in *Network5*, the resultant modularity was *0.232*. Holding the topic constant and mixing the time period in *Network6*, the modularity was *0.166*. Finally, in a single period, with a mix of topics, the modularity for *Network7* was *0.213*. These sample results can be interpreted through the pattern of topical similarity being inversely proportional to modularity. A corpus made up of articles with qualitatively different topics projected a network with a high modularity scores. Conversely, a corpus made up of articles containing qualitatively similar topics projected a concept network with low modularity scores. Another assumption in this analysis was that topics changed over time. Varying the date period was used to isolate separate discussions. However, *Network6* showed a much lower modularity score than both *Network1* and *Network 3*. This indicates that

the topics within were closely related. Although these are only one quarter apart, it suggests that topics can have continuity over time.

3.8 Conclusion

In this chapter, I outlined the method used to answer the research question about tracking a product ontology from a product public. The empirical strategy I use in this dissertation was a study in two parts. The outcome of the first part guided the sampling for the second. For Chapter 4, I looked at the structural user dynamics in a popular tech blog to see if a product public is more likely to emerge around particular sections in a web forum. This was aimed at drawing out where community and public activity occurs. This is particularly important for uncovering the emerging product ontology. Considering the results from this analysis, I was able to focus on front page posts of four major tech blogs to find the emergence of meaning structures. However, to do this I needed to develop a method for operationalising a relational product ontology. This chapter outlined how this method works and provided several examples. For Chapter 5, I used this novel method to conduct an analysis of the emergence of tablet product ontology.

I reported here on an example of computational discourse analysis of tech blogs. The tablet device market has only recently picked up and seemingly has converged around a dominant design that corresponds to a stable product ontology. The aim of this study is to find speculation discourse around a nascent product ontology. I look to tech-blogs as research sites worth following to find evidence of a *product public*. To conduct such an analysis, I first data-mined blog posts from the Internet, then analysed them with computational methods. Using techniques from the fields of computational linguistics and natural language processing (Pollach, 2012; Manning & Schütz, 1999), I look for meaning structures (Mohr, 1998) and discourse that occur

before the official introductions by firms signalling participation in a radical new product category.

Thus far, I have reviewed approaches to concept networks and developed a novel method as an extension to Carley (1993) and Corman et al., (2002). The method contributes to the projection of concept networks by overcoming weaknesses of both approaches. I used multiword noun phrases as unit level concepts within a sentence. I commensurated these into a network by collapsing sufficiently similar noun phrases into nodes. This was moderated through the three parameters to the method. The final contribution of this method is in using modularity (Leicht & Newman, 2008) to measure concept networks. As shown with several examples, modularity corresponds to topical cohesiveness within a network.

With this method, I can begin to project and measure meaning structures as concept networks. One such meaning structure will be the product ontology. In chapter 5, I apply this method over longitudinal slices of discourse to track the emergence of the meaning structure. Modularity will offer a network structural measure, whilst filtering for centralities of nodes and high edge weights will allow me to focus on varying level of association intensity. By following the different levels of intensity, I should be able to track how cognitive structures become embedded over time.

I developed four empirical propositions for the dissertation based on theory and sample data for projecting and analysing concept networks with this method. These will help to guide assumptions of the next two empirical chapters.

Proposition 1: *public and community activity are organising principles that create different shapes of discussion network structure*

Proposition 2: *given the right setup parameters, the overlap of noun phrases across sentences projects the structure for a concept network.*

Proposition 3: *a stable empirical product ontology will be a nomenclature that corresponds to a concept network held together by broker nodes that link smaller modules to form a structure with the label in the core.*

Proposition 4: *for a corpus made up of articles that are qualitatively more topically similar, the modularity scores are lower.*

Chapter 4 - The ecology of a tech blog: communities and product publics

4.1 Introduction

In this chapter, I empirically examine the organising principles of *community* and *product public* activities in a tech blog. I look for evidence of structural signatures by looking at the aggregate dynamics of thread categories. Looking at the variables of thread, posts, and users, I organise the analysis into comparing aggregates through proportions, user overlap and timing dynamics.

I address the chapter's research question of whether *product publics more likely to be found on the front stage or backstage of online discussion forums* by comparing dynamics of thread categories that correspond to the blog homepage and the user forum. This is based on the insights from Kaltenbrunner et al., (2009) and Gonzalez-Bailon et al., (2010) that suggest topics organise the norms and structures of discussions. The first proposition in chapter 3 suggested that publics and communities are organising principles that, when aggregated, will lead to different shapes of structure as discussion networks. Gonzalez-Bailon et al., (2010) showed that topic discussions adhered to patterns that corresponded to structural signatures. This insight lends itself to thinking about how a single website could host both public and community activity.

As discussed in previous chapters, this construct of the *community* has been operationalised as a core-periphery structure (Borgatti & Everett, 1999). The core is

made up of dense strong ties and the periphery is made up of weak ties. An alternative organising structure I have developed in this thesis is to treat online collectivities as *publics*. Under this, the main variable of social ties is replaced with topical ties. The structure comes from actors developing similar discourses. When a discourse stabilises into an ordered meaning, this can be seen as a relational structure. This first condition of a public does not preclude a community as being mutually exclusive, however the second and third conditions of ephemeral interaction and temporary suspension of identities run counter to the social structure of community. A public does not stabilise to form social ties, identities or reputations and this therefore a temporary activity. In chapter 2 developed the view that the structured outcome of a public is a meaning structure, like a product ontology. I drew on the Pragmatist approach has its roots in the sociology of collective behaviour, most notably in Shibutani's (1966) work on rumours. Thus, to think about a public empirically, I developed the criteria of *ephemeral interaction*, *temporary suspension of identities*, and *topical focus*.

The full *macrumors.com* dataset contains 133556 discussion threads that the site has organised into eight broad categories. These each have two levels of subcategories that make up a classification hierarchy. In this chapter, I will conduct an analysis through comparing across those categories to find evidence of publics and communities.

4.2 Descriptive measures

In this section, I show a descriptive summary of the data. This will provide the three analyses in the chapter with some context. The four main data constructs I use in this are *posts*, *threads*, *users*, and *thread categories*. Table 4.1 show the total sum of each construct in the 13 month dataset of *macrumors.com*. These data constructs are related, as each user writes a post, and these posts are each contained in threads. Each thread is assigned to a thread category. In turn, each thread category is organised into a hierarchy of categories. At the top level, there are 8 thread categories.

total posts	1460102
total threads	133556
total users	78785
thread categories	58

Table 4.1 - Description of overall activity

The distribution of posting behaviour in threads is presented in Figures 4.1a by using a bin width of 100 posts. From this, there are a significant number of threads with less than 100 posts, and a small number of threads each containing over 1500 posts. There still is a strong right skew. The cumulative distribution function in Figure 4.1b shows the right skew and that about 99% of the threads contain less than 250 posts. The mean number of posts is 10.93 and the medium is 5.

One feature of the distribution is the number of gaps at the high end. In Figure 4.1a, we see there a less than 10 threads with > 1500 posts. Within those, there are no threads with 1800-2200 posts. It is not clear why there are less than 5 threads with over 2200 posts.

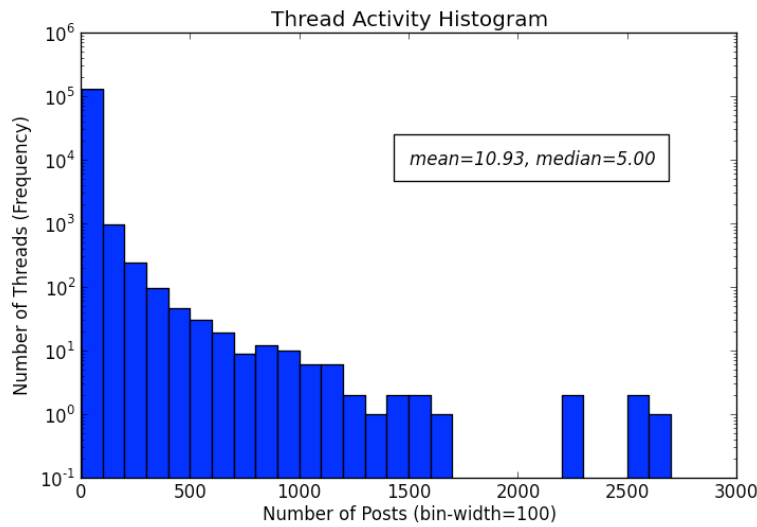


Figure 4.1a: Histogram of the number of Posts per Thread

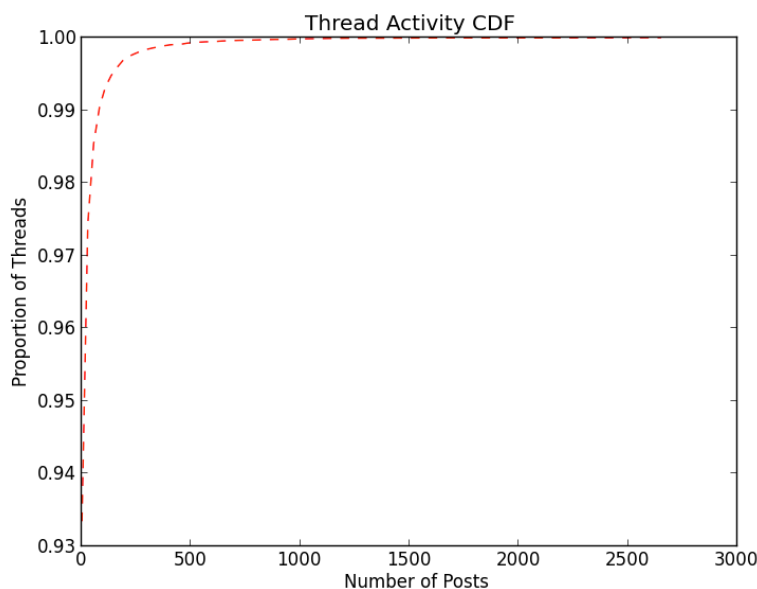


Figure 4.1b: Cumulative Distribution Function (CDF) of the Number of Posts per Thread

In Figure 4.2a, user posting activity is presented as number of posts per user. This is also right skewed, indicating that most of the users do not write many posts. The median is 2. From the CDF in Figure 4.2b, we see 99% of users posted less than 300 posts in the sample. Interestingly, there are less than ten prolific users who have post activity ranging from 4000 to almost 8500.

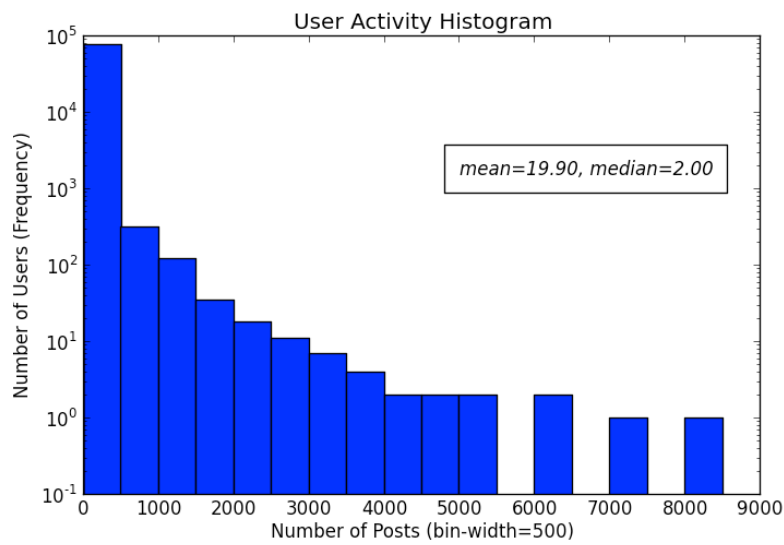


Figure 4.2a: Histogram of the number of Posts per User

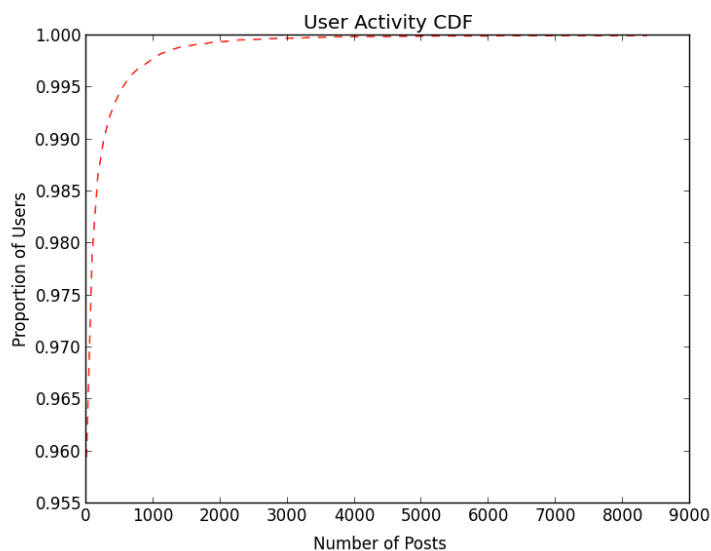


Figure 4.2b: Cumulative Distribution Function (CDF) of the Number of Posts per User

The two features of note in these distributions are right skews of posts per thread and posts per user. There are many users that seldom post and only a few who are prolific. Similarly, most threads have few posts and a several have over 1500. These both point to a core-periphery community structure (Borgatti & Everett, 1999).

Full thread category	Threads	Unique Users	Posts
Apple Applications/iPhone and iPod touch Apps/	5	123	87
Apple Applications/Mac Applications and Mac App Store/	13542	13156	68159
Apple Hardware/Apple TV and Home Theater/	1193	2923	12770
Apple Hardware/Buying Tips and Advice/	7731	9273	78011
Apple Hardware/Desktops/iMac	3659	6078	44513
Apple Hardware/Desktops/Mac mini	542	1562	5864
Apple Hardware/Desktops/Mac Pro	2533	4238	37454
Apple Hardware/Mac Peripherals/	4328	5895	23910
Apple Hardware/Notebooks/MacBook	12	438	160
Apple Hardware/Notebooks/MacBook Air	203	2019	8140
Apple Hardware/Notebooks/MacBook Pro	14326	14913	149971
Apple Hardware/PowerPC Macs/	2531	4034	16971
Apple Systems and Services/Mac Basics and Help/	10099	9247	57148
Apple Systems and Services/Mac OS X/	11267	13473	84281
Apple Systems and Services/Mac OS X/Mac OS X 10.7 Lion	2	88	96
Apple Systems and Services/Mac OS X/Mac OS X Server, Xserve, and Networking	52	141	189
Apple Systems and Services/Programming/Mac Programming	1734	1772	11650
Apple Systems and Services/Windows on the Mac/	3036	4317	16817
Archive/Archives of Old Posts/Event Archives	112	884	2274
Archive/Archives of Old Posts/Event Archives/Macworld San Francisco 2008	215	2029	5462
Archive/Archives of Old Posts/Games	1046	1390	16928
Archive/Archives of Old Posts/iPhone Purchaser Meetups	136	1586	5034
Archive/Archives of Old Posts/Leopard Event Meetups	168	1161	2498
Archive/Archives of Old Posts/Marketplace Archive 2	5319	2685	31953
Archive/Archives of Old Posts/Marketplace Archive 2/iPhone Marketplace Archive	49	140	326
Archive/Archives of Old Posts/Music Discussion	448	1857	6368
Archive/Archives of Old Posts/Past Contests/4,000,000 Post Contest	19	130	227
Archive/Wasteland/	792	2569	10030
iPhone, iPod and iPad/iPhone/	8919	13273	106276
iPhone, iPod and iPad/iPhone/iPhone Accessories	639	2487	5887
iPhone, iPod and iPad/iPhone/iPhone Hacks	2101	4469	16757
iPhone, iPod and iPad/iPhone/iPhone Tips, Help and Troubleshooting	3194	7926	19844
iPhone, iPod and iPad/iPod touch/	4266	5478	40120
iPhone, iPod and iPad/iPod touch/iPod touch Accessories	296	1048	2587
iPhone, iPod and iPad/iPod touch/iPod touch Hacks	1433	2201	11977
iPhone, iPod and iPad/iPod/	4352	5865	28748
Mac Community/Community Discussion/	4641	5463	102311
Mac Community/Community Discussion/Apple, Industry and Internet Discussion	1246	2995	14462
Mac Community/Community Discussion/Current Events	588	1804	13941
Mac Community/Community Discussion/Politics, Religion, Social Issues	1323	1323	51024
Mac Community/Picture Gallery/	374	4497	18873
Mac Community/Site and Forum Feedback/	623	1258	8818
Mac Community/Site and Forum Feedback/Mac Guides	16	116	131
News and Article Discussion/DailyTunes Song Discussion/	312	362	962
News and Article Discussion/iOS Blog Discussion/	233	3718	11048
News and Article Discussion/Mac Blog Discussion/	209	4238	13972
News and Article Discussion/MacBytes.com News Discussion/	2197	2887	16421
News and Article Discussion/MacPolls.com Discussion/	41	2146	2558
News and Article Discussion/MacRumors.com News Discussion/	982	16227	175847
Special Interests/Apple Collectors/	194	537	1980
Special Interests/Digital Audio/	880	1350	4776
Special Interests/Distributed Computing/	39	112	354
Special Interests/Mac and PC Games/	1489	4013	15019
Special Interests/Mac and PC Games/Console Games	922	1409	18583
Special Interests/Visual Media/Design and Graphics	1118	1897	9789
Special Interests/Visual Media/Digital Photography	1908	2634	23829
Special Interests/Visual Media/Digital Video	2332	2897	14481
Special Interests/Visual Media/Web Design and Development	1588	1976	11426

Table 4.2 - Thread categories and thread frequencies, total users, posts

In Table 4.2, the thread categories are outlined with descriptive frequencies of threads, users, and posts. The categories are arranged in a hierarchical ordering with eight at the top-level: *Apple Applications*, *Apple Hardware*, *Apple Systems and Services*, *Archive*, *Mac Community*, *News and Article Discussion*, *Special Interests*, *iPhone*, *iPod* and *iPad*. There are 31 second-level thread categories and 34 at the lowest level.

I constructed this table to show the immediate category each thread would be placed when it was created. Because this data collection was done in 2011, some of the categories have been reshaped since 2007, when these posts were originally written. One example of this is the top level category *iPhone*, *iPod* and *iPad*. Back in 2007, the iPad had not been announced yet, so this category would have been titled *iPhone*, *iPod*. There is a small risk that the post-2007 revision of these categories might affect the analysis in 2011. However, a sensitivity analysis done on the original date-stamps on the posts indicated that the only threads captured in the dataset were started between Jan 2007 to Jan 2008.

The thread, user, and post counts were included in the table to indicate how the activity was dispersed across the thread categories. This will be organised and analysed further in subsequent sections. One interesting aspect of these counts is shown in Table 4.3 with simple linear correlations. Threads and User counts were highly correlated at 0.83 . This suggests that overall, the length of a thread is partially determined by the number of users. However, with this crude measure, this is not a causal argument. Users and Post counts were also highly correlated at 0.89 .

However, Threads & Posts were correlated lower at *0.70*. This is a strange result that will require further analysis to explain.

	Correlations
Threads & Users	0.82
Threads & Posts	0.70
Users & Posts	0.89

Table 4.3 - Correlations between variables frequencies in thread categories

With Table 4.4, I looked more closely at post counts. This table lays out the top eight thread categories as organised in descending order by post percentages, normalised against the total. The highest number of posts are in the category *News and Article Discussion/MacRumors.com News Discussion/*. This category also has 21 percent of the users. However, these threads only represent 1 percent of all those on the site. This is interesting in comparison to *Apple Hardware/Notebooks/MacBook Pro*, which has a relatively proportionate 10 percent of the posts written by 19 percent of the users in 11 percent of the threads. These categories are beginning to show structural signatures that appear in disproportionate distributions across threads, posts and users. This leads to the first step of analysis.

Thread category	Thread frequency	Thread frequency %	Post Counts	Post Count %	Unique Users	Unique Users posting %
News and Article Discussion/ MacRumors.com News Discussion/	982	0.01	175847	0.12	16227	0.21
Apple Hardware/Notebooks/MacBook Pro	14326	0.11	149971	0.10	14913	0.19
iPhone, iPod and iPad/iPhone/	8919	0.07	106276	0.07	13273	0.17
Mac Community/Community Discussion/	4641	0.03	102311	0.07	5463	0.07
Apple Systems and Services/Mac OS X/	11267	0.08	84281	0.06	13473	0.17
Apple Hardware/Buying Tips and Advice/	7731	0.06	78011	0.05	9273	0.12
Apple Applications/Mac Applications and Mac App Store/	13542	0.10	68159	0.05	13156	0.17

Table 4.4 - Post, Thread and User activity in top Thread sub-categories

4.3 Proportions of activity

Thus far in the chapter, I have described the dataset of posts, threads, and users through their distributions and some aggregate activity in thread categories. The distributions have provided some evidence of community structure. Looking across thread categories showed that activity was not homogenous across thread categories. This section will look closer at the proportions of posts, threads, and users for further evidence of community or public activity.

The main focus of this chapter is around discursive behaviour relating to structural dynamics. This leads to a question about characterisation of strong community ties or temporary discursive groups. I use thread category data to partition the structural measures and characterise behavioural activity. There are different hierarchical categories that classify threads according to topics. I use the insight from Gonzalez-Bailon et al., (2010) as a starting assumption about threads within a category. This then simplifies the analysis to look across thread categories using the variables of post, thread, and user activity. I look both at the thread category and the aggregate thread categories at the top level.

In observing the aggregate profile of users who participate in specific sections and whether there are substantially different populations or dynamics across them, I begin to directly address this chapter's research question. The assumption I use here is that topics are contexts for discussion networks. They have mechanisms which lead to structural forms. This is in line with previous research in new structuralism (Lounsbury & Ventresca, 2003; Mohr, 1998; Gonzalez-Bailon et al.,

2010). Thus, it becomes of interest to compare dynamics across categories of threads.

In Figure 4.3, the thread subcategories from Table 4.2 are laid out on the x-axis in descending order of post counts. Then, normalised post counts and unique user counts are plotted for each subcategory. These two variables were compared in Table 4.1 and showed a 0.89 correlation. This shows that there is an overall similar trend in several subcategories between the users who get involved and their tendency to post. However, in looking at this chart, there is substantial variation of user involvement across different categories. Hence, we can infer from this chart that users are not homogenous in their activity across thread categories.

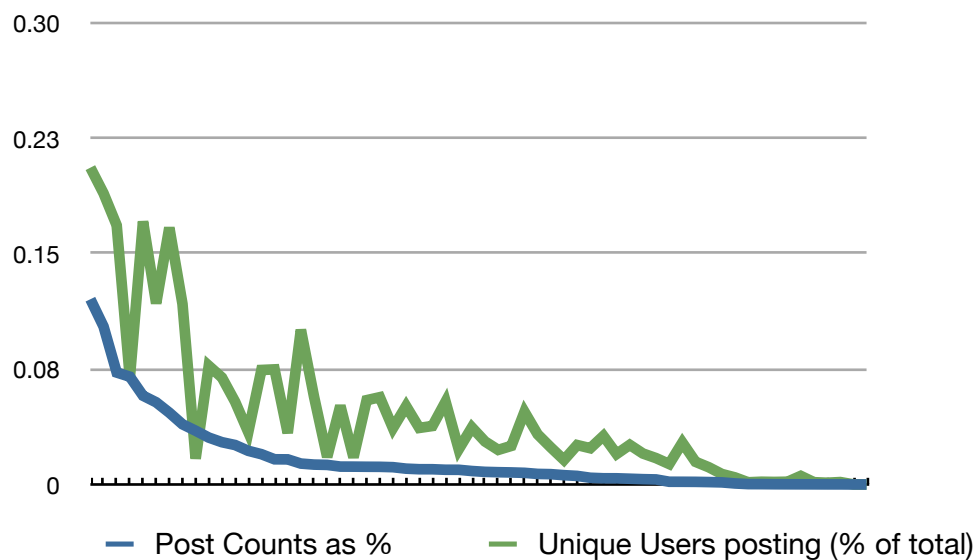


Figure 4.3: Normalised percentages of thread sub-categories, Post counts and Unique User participation

Figure 4.4 and Table 4.4 explore this variation in activity across thread categories in further detail. By plotting the top seven subcategories by percentage of overall-posts on the site, Figure 4.9 shows that *News and Article Discussion/MacRumors.com News Discussion/* and *Apple Hardware/Notebooks/MacBook Pro* contain the highest number of posts amongst the categories. From Table 4.4, the former not only has 12% of all posts from the sample, but also the involvement of 21% of the total users. Yet, this *News and Article Discussion/MacRumors.com News Discussion/* category only accounts for 1% of the threads on the site. In the *Apple Hardware/Notebooks/MacBook Pro* category, it has a more balanced 11% of the threads, accounting for 10% of all posts, but involving 19% of users on the site. When we compare these to the *Mac Community/Community Discussion/* category, a more balanced picture emerges. It accounts for 3% of the threads, 7% of the posts and involves 7% of the users. Other categories display posting activity variation as well.

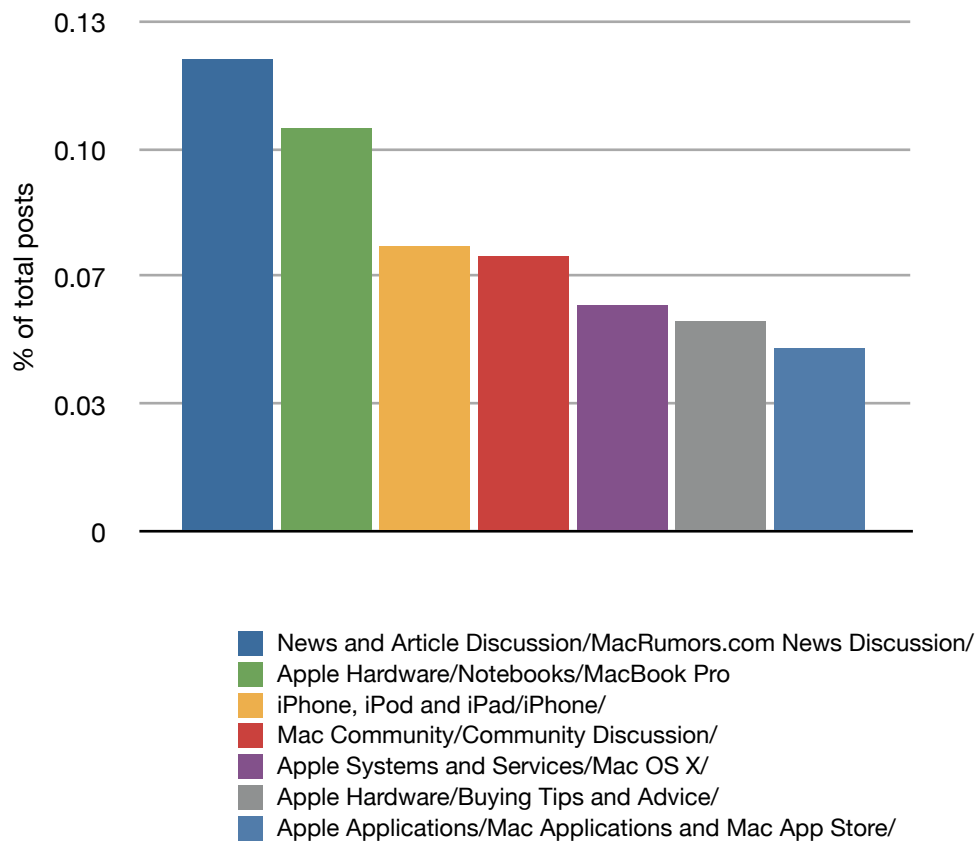


Figure 4.4: Sub-categories containing largest percentage of posts

The subcategories of *News and Article Discussion/MacRumors.com News Discussion/* and *Mac Community/Community Discussion/* are compared in Table 4.5. In Table 4.5, there is evidence to support the idea of public-like activity being associated with the *Macrumors.com news discussion* category and community-like activity being associated with the *Community discussion* category. Next to the average number of users as 3578, the news category has a high number of users at 16227, and a low thread count of 982. The average thread count across the categories is 2226. Conversely, the community category has lower users 5463 and

higher threads 4641. Users in the community category participate in an average of 11.8 different threads, whereas the news category only has 6.2.

	News and Article Discussion/ MacRumors.com News Discussion	Mac Community/ Community Discussion
Users - unique	16227	5463
Threads	982	4641
Posts	175847	102311
Users: thread involvement mean	6.184	11.8
Users: thread involvement stdev	18.95	38.32

Table 4.5 - Comparing two thread sub-categories

At this point in the analysis, I see some evidence to show that two thread categories may correspond to public and community activity. However, these are both second level categories and together only represent less than 20 percent of the posts on the site. To get a bigger picture of how activity varies across the topics on the site, I aggregate the subcategories into their parent categories and continue the analysis.

4.3.1 Thread categories: aggregates

Table 4.6 reports on the exercise of aggregating subcategories into their parent categories and the corresponding grouped measures. Here, the dispersion of activity is shown across threads, posts, and unique users. What is interesting here as a first analytical point is the ranking of activity. The top categories for percentage of threads

are: *Apple Hardware; Apple Systems and Services; iPhone, iPod and iPad*. The top categories for percentage of posts are *Apple Hardware; Mac Community; News and Article Discussion; iPhone, iPod and iPad*. With this dispersion, there is reason to think that activity changes across topical categories.

Top level category	Thread frequency in sample	Thread frequency %	Post Counts	Post Counts as %	Unique Users	Unique Users as %
Apple Applications	13547	0.10	68246	0.05	13229	0.17
Apple Hardware	37058	0.28	377764	0.26	29125	0.37
Apple Systems and Services	26190	0.20	170181	0.12	21903	0.28
Archive	8306	0.06	81110	0.06	9375	0.12
Mac Community	8811	0.07	209560	0.14	9931	0.13
News and Article Discussion	3974	0.03	220808	0.15	19996	0.25
Special Interests	10470	0.08	100237	0.07	11331	0.14
iPhone, iPod and iPad	25200	0.19	232196	0.16	27904	0.35

Table 4.6: Aggregate thread categories: thread and posting activity

Figure 4.5 shows thread and post frequency percentage across the aggregate thread categories. This demonstrates that most thread categories have similar thread percentage and post percentage measures. However two categories that stand out from this chart are *Mac Community* and *News and Article Discussion*. Both of these have a greater percentage of posts than threads. Furthermore, in the comparison between these two categories, the user percentage measure shows that 0.25 of the site is in *News and Article Discussion*, but only half of that is involved in *Mac Community*. Why do these two categories have a disproportionate amount of posts relative to threads? This is an interesting question given the idea of a community

structure having a dense core and a light periphery. *News and Article Discussion* has both a relatively greater percentage of posts and a lower percentage of threads. With a larger user percentage as well, this would suggest a core density of activity. However, both of these are eclipsed by the *Apple Hardware* aggregate category.

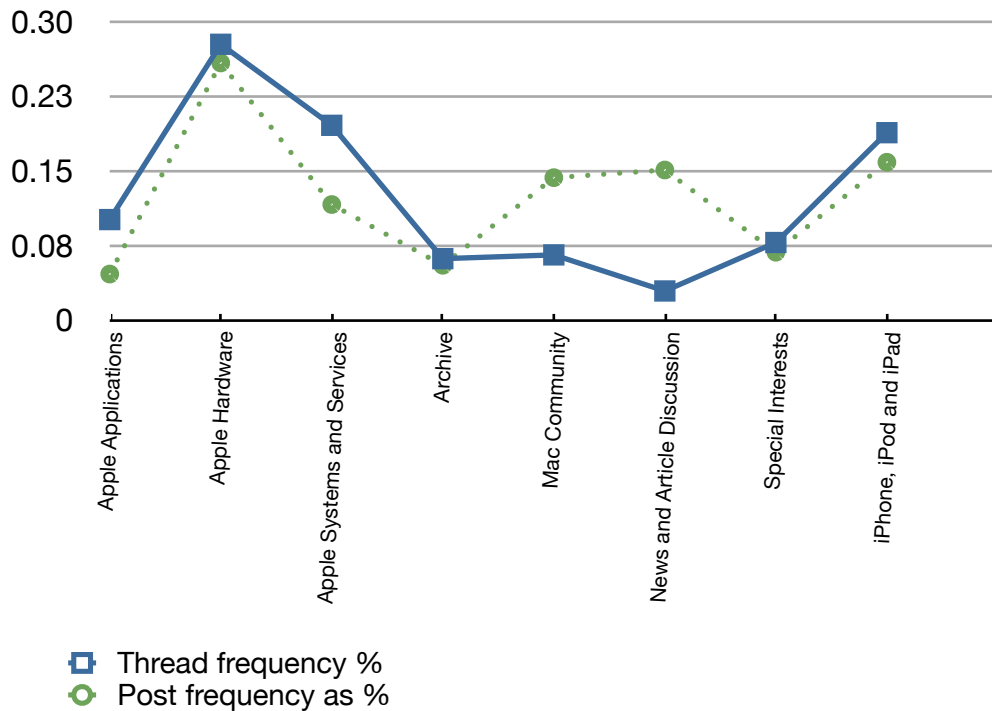


Figure 4.5: Thread and Post counts across Thread Categories

Another point that comes from Figure 4.5 is that *Apple Hardware* is a substantive aggregate category, representing the largest number of threads, posts, and users in the sample. Subcategories indicate these conversations are organised around product specific talk. Table 4.2 showed the largest individual thread category within the aggregate to be *Apple Hardware/Buying Tips and Advice/*. This characterises the activity in the aggregate category as potentially being about advice around buying

decisions. Given that the sub categories are based on existing products, the aggregate category also be characterised as being about usage discussions. From this evidence of proportions, we see that three aggregate categories are notably different. Just looking at proportions was a crude form of analysis, but it did show some interesting descriptive usage statistics. Two further analyses to show structural dynamics will be user overlap and timing dynamics.

The core question in this chapter is one around characterising activity on this site as being community-based or a more ephemeral discursively linked public. On the surface, the categories of *Community Discussion* and *Macrumors.com news discussion* categories seem to be good proxies for community and public, respectively. The former is somewhat self-explanatory. The latter could be considered a discursive space where users temporarily come together to deliberate on particular topics. I would expect issue public-like activity to be episodic, perhaps with bursts of activity in fewer threads. Whereas, I would expect community activity to be characterised by more balanced posting activity spread across many threads. Community activity could include goodwill type generalised exchange (Faraj & Johnson, 2010) where users are helping one another out. Examples could include IT support, or free-revealing user-innovation solutions (Franke & Shah, 2003). Furthermore, if the organising principle is one of community, then users themselves are likely to be spreading their activity around categories. Whereas, a public like organising principle would suggest users get involved temporarily in focused areas and do not stay to form any bonds or get involved in any goodwill or support activities.

4.4 User Overlap

In Chapter 3, I outlined the methodology used by Kaltenbrunner et al., (2009) to trace user activity overlap. This is an important test in tracing where particular users are participating and to what degree. The method for user overlap for this chapter first constructed user participation vectors, and then used cosine similarity on each pair of vectors to derive a similarity measure. This technique is adapted from the TF-IDF cosine similarity method that is often used in information retrieval to compare texts in a corpus. Kaltenbrunner et al., (2009) provided the insight that this could be used to measure user participation in discussion networks. By constructing user participation vectors for aggregate thread categories, this enables an analysis of broader users to see where they are interacting. In this section, I report on the user overlap analysis done across the aggregate thread categories. This will provide a profile of user activity on the site.

In Figure 4.6, the effect of user activity overlap across categories is shown through a shaded matrix. The rows and columns represent categories and dark shades indicate more similar user activity and lighter ones. By definition, the same category compared to itself will have perfect similarity.

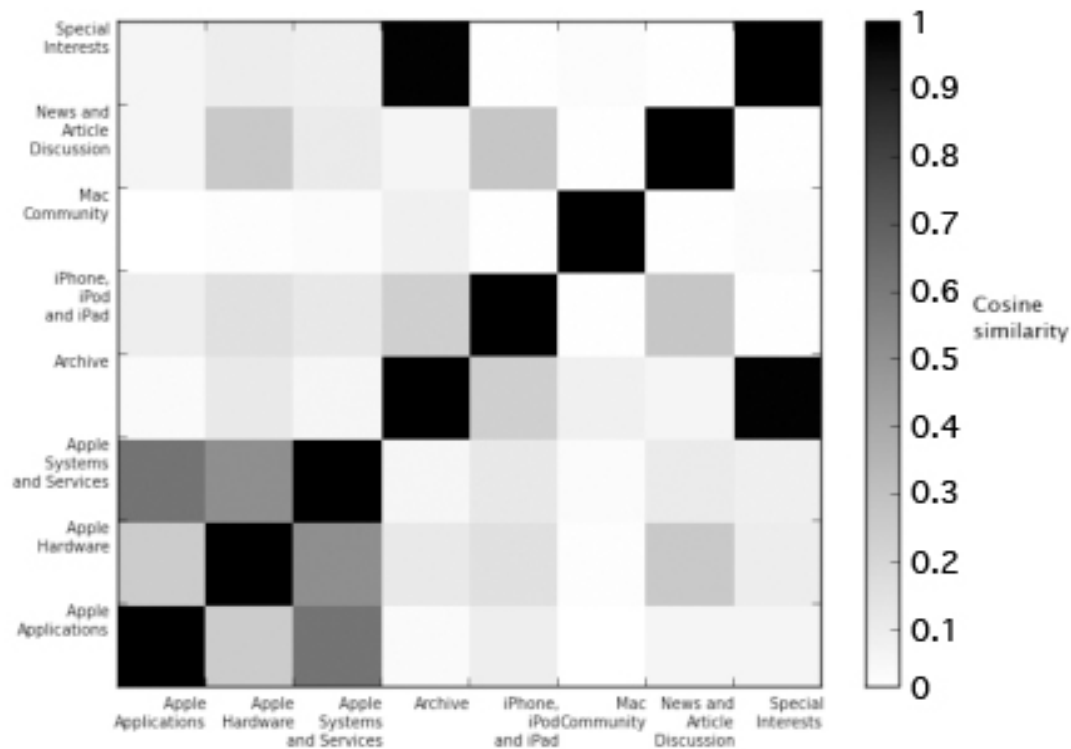


Figure 4.6: User activity overlap across thread categories

In this figure, there is a clear cluster along the bottom left corner between *Apple Applications*, *Apple Hardware*, and *Apple Systems and Services*. This indicates that the same users, in similar numbers of posts are frequenting those three aggregate categories. Interestingly, *Mac Community* does not seem to be very similar to other categories. This would indicate that users participating here are not participating in other categories. Because this measure indicates both the presence of users and their activity in each aggregate thread category, there is a possibility that the same users are participating but at drastically different levels of activity.

However, here the similarity measures for *Mac Community* to others are at an extreme values of below 10 percent, so this can be ruled out.

Users who post across overlapping categories provides evidence of community like organising principles. Conversely, if users only participate in one section of the site,

this raises questions about how they enter the site and why they do not stay around.

Figure 4.6 shows a matrix of overlapping user activity according to profiles. By looking at a matrix of user activity, this evidence that users cluster across particular categories. Another way to use a useful way to visualise this clustering is through a dendrogram.

User participation profiles can be used to find clusters of similar activity by constructing a dendrogram. In such a diagram, clustering is shown from the bottom up, with lower levels corresponding to tighter clusters. For this data, similar thread categories provide evidence of groups of users participating in similar ways. Figure 4.7 is a dendrogram that shows how this data forms into clusters. In this dendrogram, there are seven levels of clustering that each correspond to a grouping. At the highest level, all the components are a single group.

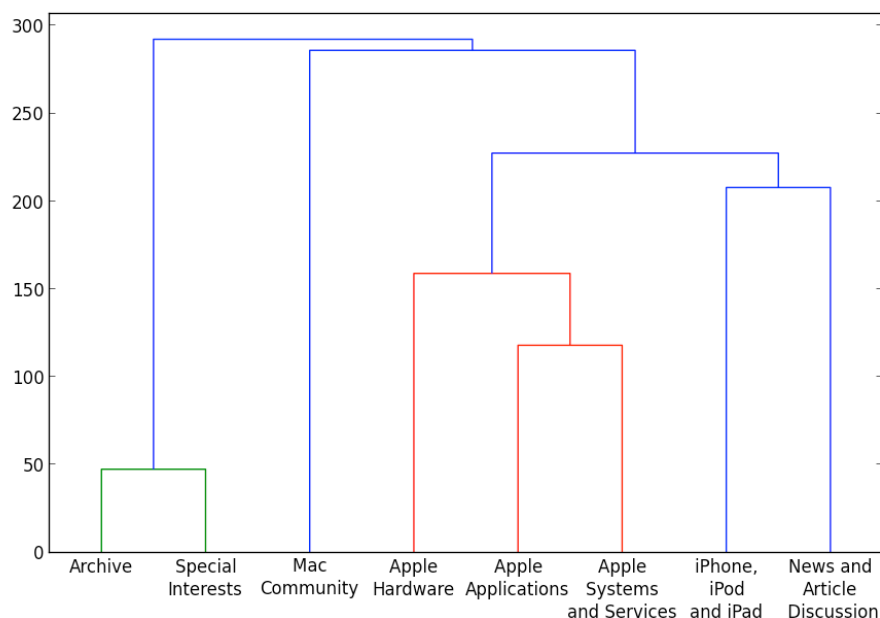


Figure 4.7: Dendrogram showing user activity overlap in thread categories

On Figure 4.7, one cluster of similar categories (highlighted in red) is that of *Apple Applications, Apple Hardware, Apple Systems and Services*. Previous evidence in this chapter indicates that these three categories take on the characterisation of Apple support forums, where users ask for help with their technology and other users respond with answers. This also supports the suggestion in the previous section about users sharing buying advice. The fact that these categories cluster in such a way suggests this may indeed be the case. This activity provides evidence of community activity.

Another cluster of note brings together *iPhone, iPod, and iPad* with *News and Article Discussion*. Given that this grouping clusters with other categories at a higher level, we can infer that the users who frequent these threads do not tend to participate in other sections on the site. They do, however, cluster with one another at a fairly high level. This may be due to a connection over the tablet and iPhone speculation. The date range corresponds to the release of the iPhone and to rumour and speculation around an Apple tablet. Given the skewed proportions of *News and Article Discussion* from Table 4.6, it seems strange that 25 percent of the users comment on 15 percent of the posts, but do not participate much into other aggregate thread categories. This clustering provides evidence to support public-like activity, where users come in and discuss rumours and speculation about future products. One final test will be to look at timing dynamics.

4.5 Timing dynamics of aggregate thread categories

In analysing the dynamics to find community and public activity, timing is important.

A public has been theorised as a temporary discursive grouping of *ephemeral interaction*. In contrast, a community structure be characterised by more stability with steady activity due to strong social ties. This distinction can be operationalised here though the temporal dynamics of posting activity.

As a baseline timing distribution, Figure 4.8 shows a cumulative density (CDF) for the entire site sample. This is a steep curve and demonstrates the long tail of timing on posting activity. The majority of posts are written quickly after the thread is initiated. In the aggregate 90 percent of all posts come within the first 1000 hours of threads starting, then the curve flattens out and remaining posts stream in more distanced.

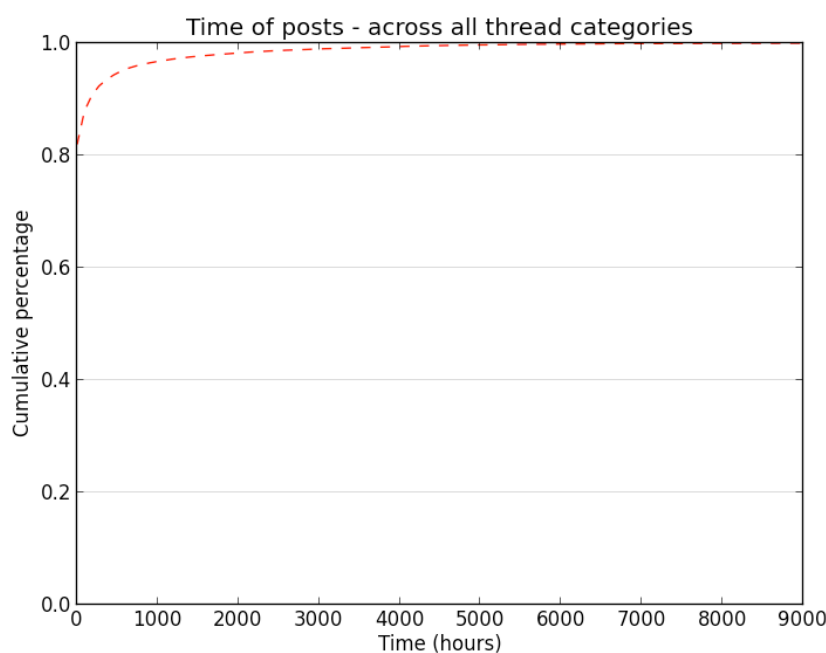


Figure 4.8: Time of postings in all threads (initial post at time = 0)

In Figure 4.9, distributions of post timing are shown across the aggregate categories in box and whisker plots. These timing dynamics have been looked at across the time spans of a year, a month and a week. The Figure is split into subfigures showing the respective time spans from top to bottom. This is blown up to show the extreme details of each time scope. Due to the long flat tails of each distribution, the year timespan shows a full picture. The bottom figure shows the quartiles representing the 25th and 75th percentiles for each aggregate thread category. Looking across each subfigure, there is a clear trend of the *News and Article Discussion* category having a relatively small tail, with the 75th percentile appearing in less than a single day. This indicates that most of the threads within this category are short-lived and have posts that come in very quickly. Comparing this to the other categories, we see longer ranges for both 75th percentiles and the tails. The *community* category exhibits a tail that extends well into 350 days and has the 75th percentile of posts coming in close to four days since threads were initiated.

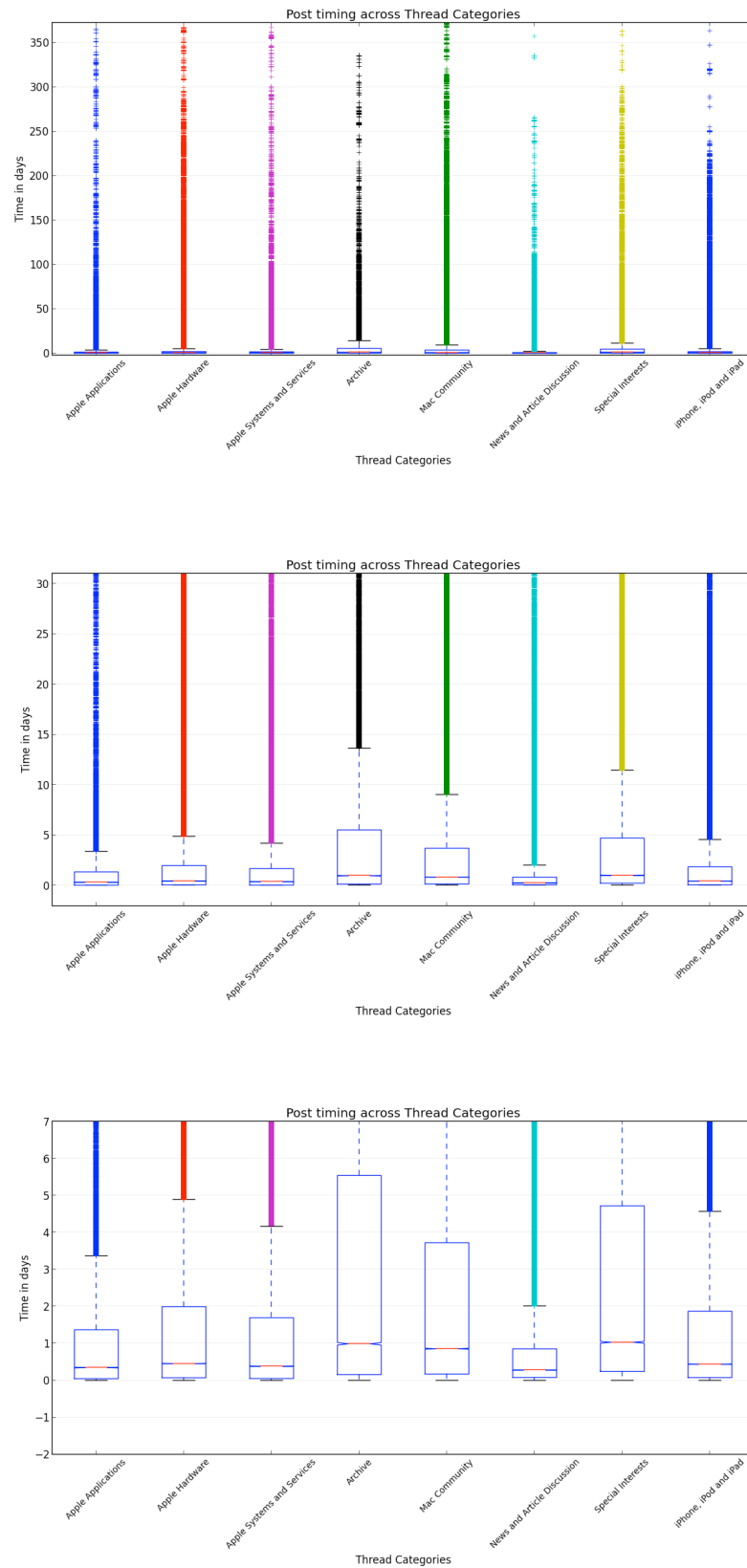


Figure 4.9: Box plots of post timing across thread categories: as year, month, week

4.5.1 Life-cycles of threads

The previous section demonstrated that activity in the *News and Article Discussion* aggregate thread category is characterised by bursts of posting activity and relatively low variation. To expand on this analysis, I consider here the lifecycle of each thread as a curve, plotted on a date line. In Figure 4.10, all threads with more than 200 posts have been plotted to demonstrate an overall picture of site activity. Here, clearly some threads burst up as intense and short spans of activity.

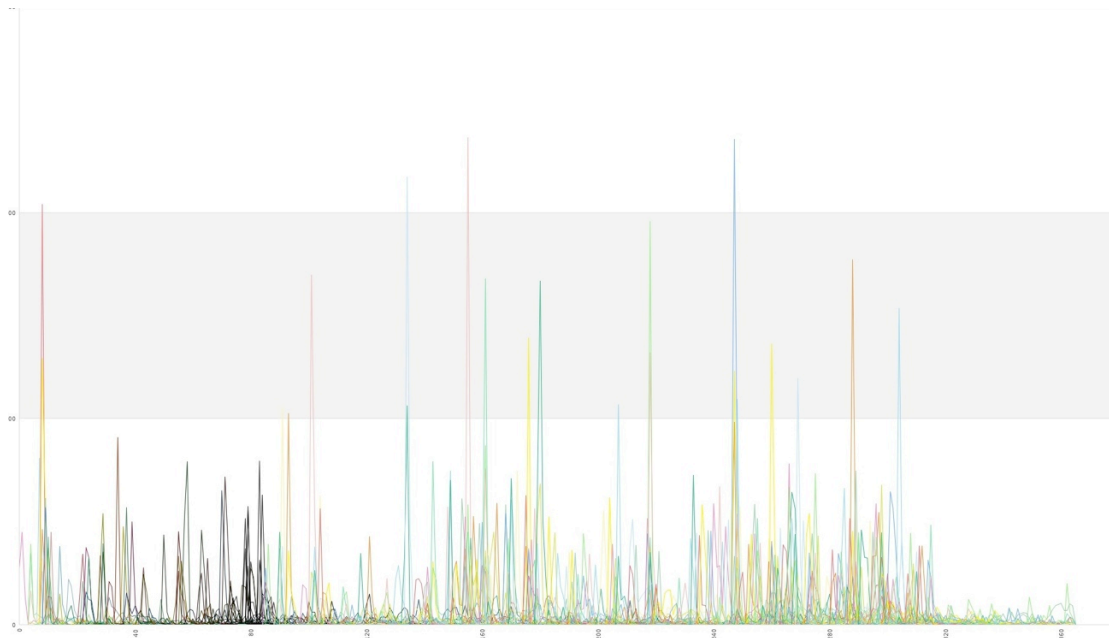


Figure 4.10: All threads with > 200 posts; lifespans as posts per day

An analysis of Figure 4.10 is that of Table 4.7. This considers the spiked dynamic from the statistical measures of skewness and kurtosis. Here, skewness is a measure of asymmetry, or the degree to which a distribution presents a long tail. The higher the skewness, the more a tail falls to the right and is longer than the left. The other measure is that of kurtosis, which measures peakedness or flatness relative to

a normal distribution. For this test, I mapped out the shapes of the thread life-cycles in each thread category.

	Mean Skew	Mean Kurtosis
Apple Applications	0.07	-2.44
Apple Hardware	0.09	-2.31
Apple Systems and Services	0.08	-2.40
Archive	0.20	-1.95
Mac Community	0.25	-1.81
News and Article Discussion	0.57	-0.98
Special Interests	0.18	-1.98
iPhone, iPod and iPad	0.11	-2.28

Table 4.7: Measuring skew and kurtosis of thread-lifespans across thread categories

Across all categories, the *average mean skew* = 0.194, and the *average mean kurtosis* = -2.02. On both metrics, *News and Article Discussion* contain the maximum values in the table. An average skew for threads in this category > 0 means that there is a tendency for right skews. An average kurtosis of -0.98 indicates that the distribution is flatter than a normal distribution (=0). However, given that it is higher than other categories this means the category has more shape to it (and is less flat). This corresponds to an explanation of the “burstiness” of threads in this category. Recall that the average length of threads in this category has a low variance; which is consistent with the characterisation that these threads are dominated by many users coming in and discussion issues for a short period of time.

4.6 Analysis

In this chapter, I have presented data from the *macrumors.com* discussion forum for 13 months. The analysis has focused on posting dynamics to reveal structural signatures of types of usage activity. These have covered cyclical dynamics, thread posting dynamics, and participation across thread categories. In doing this, I was able to build up structural profiles of user participation and analyse how they linked to topical areas.

There are three important findings that come out of this analysis. The first is a clear picture of the important variations of what threads look like, and how they occur. This variation is pronounced around thread categories. In particular, the ‘news’ category stands out and helps to make more precise the dynamic of a *product public*, which is noticeably different relative to the general community category. Second, there are some important similarities across categories. This analysis informs three clusters of activity across thread categories that can be roughly discussed as a typology. Third, all thread categories on the forum share dynamics in that for most posts, up to the 75th percentile come in less than a week after the thread begins.

In constructing a typology of thread categories, the first is a grouping makes sense from the frame of a support forum. This includes *Apple Hardware*, *Apple Applications*, *Apple Systems and Services*. They collectively represent 58 percent of all threads in the forum and 42 percent of the all posts on the site. These together could logically be construed as somewhat of a support forum. Users start many threads asking questions about using their Apple computers. Many of the threads get

responses within 1.5 days, though some stretch on well past that. The evidence presented supports activity we would expect to see in a community.

In this second category, the *Mac Community* and *Special Interests* thread subcategories are similar. Although they do not share many overlapping users, their timing dynamics resemble one another. Both categories have the longest response times on the forum, with 75th percentile measures at just over 5 days. This indicates that users involved in these are not so rushed and are adding to threads in a relatively relaxed manner. With posting dynamics, each category has close to 8 percent of the threads and close to 14 percent of the users. Interestingly, despite having similar normalised numbers of threads, *Mac Community* has 14 percent of posts, whilst *Special Interests* has 7 percent.

The final category in this typology is that of a *product public*. The constituent thread category is mainly *News and Article Discussion*. It has 15 percent of the posts and 25 percent of users participating. Interestingly, it has only 3 percent of the threads. Combined with data on timing dynamics, we see that these threads tend to be spiky and die off quickly. The 75th percentile of posting times is 1 day and is the shortest in the sample. However, the *iPhone*, *iPod*, and *iPad* thread-category most resembles the first element of this typology, with 19 percent of threads, 16 percent of posts, 35 percent of users participating, and a post timing 75th percentile of almost 2 days with a long tail of outliers. It also shares user overlap with the 'support forum' constituents. Therefore, this cluster is a hybrid and warrants further study.

Here, the bursty threads are populated by users whom are not regular contributors across the forums. They come into to discuss an issue, then leave. Most of these spiky threads are within the *News and Article Discussion* category. However, it would be a mischaracterisation to suggest that the site is primarily for news discussion. There is also an active community that seems to prop up the platform as an active place of discussion.

4.7 Discussion

Recent papers in *Organisational Science* suggest that online communities need to be more fully and adequately theorised and empirically examined (Faraj et al., 2011; Faraj & Johnson, 2010). Faraj et al., (2011) argued that rather than being treated as static structures, online communities need to be considered as more dynamic collectives of activity, constantly in flux. They suggested that a definition of online community was difficult to stabilise, given the loose nature of these collectivities. I consider that if members are not necessarily sharing identities, or meeting up in person, then perhaps the construct of community is being misspecified with these empirical settings.

In much innovation work, the prevailing approach to date has been to adopt the community as the organising construct (Dahlander and Frederickson, 2011; Franke and Shah, 2003; Fuller et al., 2006). For example, Muniz et al., (2006) study 'brand communities' as potentially representing a valuable source of innovation for the firm. In these studies, community is core-periphery structures composed of a dense core of dedicated members who form strong, well-formed ties and maintain them through regular interaction. Borgatti & Everett (2000) argued that these structures enable superior information flows at the core. The other side of this argument is that new information is often seen as illegitimate by the core and is kept out at the periphery (Schilling, 2005).

In this chapter, I presented evidence showing both public and community activity according to thread categories. The threads within the *News and Article Discussion*

category are particularly interesting as they also appear on the front-pages of *macrumors.com*. That is, they are comment threads that are initiated by Blog editors. Therefore, it is reasonable to consider product-public type activity as occurring on the “front stage” (Faraj et al., 2011; Goffman, 1959) of the site. The next step will be looking for evidence of meaning structures within *front stage* discourse. This will be done in next chapter, using the methodology developed for analysing concept networks.

Chapter 5: The emergence of product ontologies

5.1 Introduction

Thus far, I have done an empirical analysis to show public like activity occurring on the front stage of a major tech blog. In this chapter, I generalise from this and look at the front page articles from all four tech blogs in the sample. I focus on the question of the emergence of product ontologies. This is empirically operationalising the product ontology in the relational view (Ruef, 1999) using the my novel method concept networks method derived from technology blog discourse that span several years before the settling of the recent *tablet computer* market.

The emergence of product markets have been theorised from several perspectives (Rosa et al., 1999; Anderson & Tushman, 1990), but this remains a relatively under-explored topic. The multiple dimensions that make up a nascent product market and the difficulty collecting ephemeral data (Forbes & Kirsch, 2011) make for a dynamic arena that is difficult to theorise, let alone follow empirically. As Shibutani (1986) and Wang (2009) point out, innovation has often been considered in material terms.

Given that many of the claims in markets are accepted as natural, it is important to consider how they come to be defined and shared. Porac et al., (2002) proposed that markets could be understood as four types of shared belief systems: *industry recipes, definitions of market structures, organisational reputations and product ontologies*. Of these, the latter belief of a product ontology is critical as a market core, a platform upon which others stand. Commensurability of these beliefs systems is only possible when there is a meaningful linkage bringing entities and objects

together. Therefore, it is the understanding of the genesis of the product ontology that is crucial towards our knowledge about market emergence.

In the settled state of a market, shared interfaces stabilise to enable economic analyses based on commensurable performance metrics and sociological analyses based on social network analysis of the field. Both of these perspectives presume a product ontology as a taken for granted assumption. The dominant design view from the technology life-cycle (Anderson & Tushman, 1990) suggests that there is consensus around a single design, but it does not speak to how stable the boundary conditions must be. It uses the logic of evolutionary economics to say that consumers select a design by purchasing it and other producers mimic this towards economies of scale (Geroski, 2003). It is then of interest in studying an empirical product ontology to ask questions about stability and consensus. In this, I will also explore the secondary research question, *how will a meaning structure stabilise relative to a field configuring event?*

5.3 Results

5.3.1 Concept network projections as longitudinal data

In Chapter 3, I developed the background of how the data was collected and analysed. This chapter covered a the novel methodology used for projecting concept networks from discourse. To analyse the emergence of a product ontology in the 5 year sample, I divided up the corpus into twenty different quarterly date periods.

From Figure 5.1, we see that post activity around “tablet” remained at a fairly similar level from 2007 until the end of 2008. From then until 2010 it rose sharply as Apple and Microsoft had tablet-related Innovation Disclosure Events in 2010-Q1. Thus, I will focus in this first section on three cases in the fourth quarter of each year from 2008 to 2010.

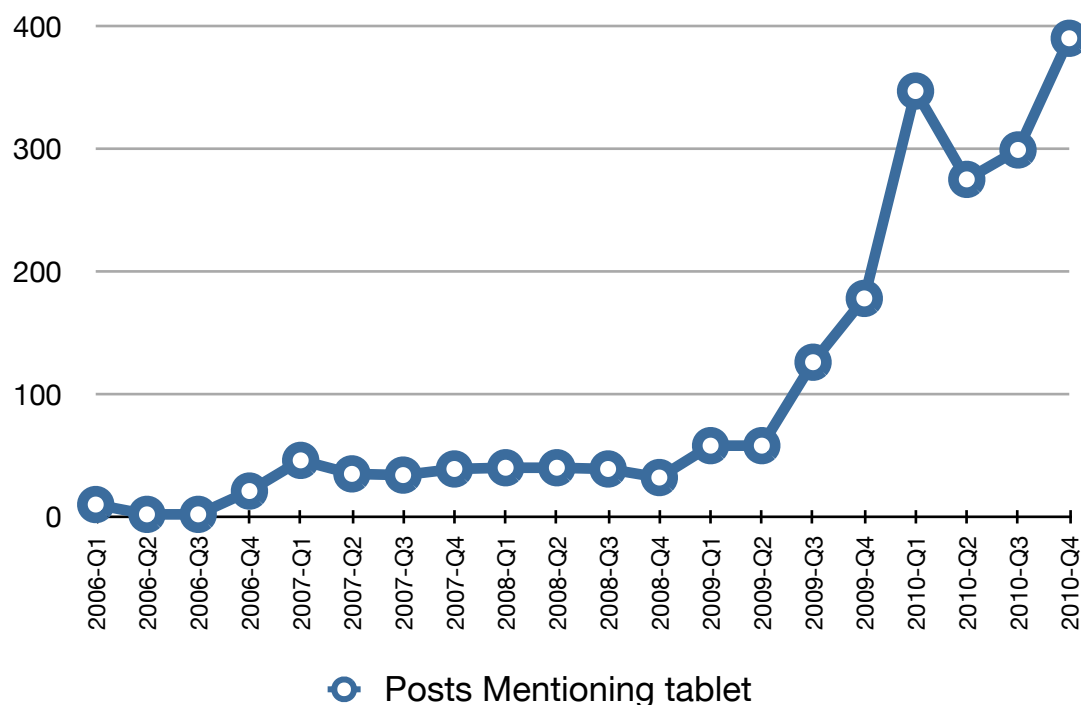


Figure 5.1: Posts containing keyword *tablet

For all network projections, parameters used in the algorithm were *thresholdsimilarity=0.9* and *degreePruneValue=5*. This means that noun-phrases with a 0.9 similarity score were collapsed into a single node and all nodes with less than 5 edges were dropped from the network. Each visualisation also has been partitioned using a community detection algorithm and coloured accordingly. Nodes and their labels are sized relative to degree value.

centre of the network, when we move to the right, there is a prominent node labelled “device.” This is connecting two clusters coloured green and magenta; the former contains netbook in the core and the latter has Apple in the core. At this level of analysis, it is interesting to consider these clusters, their central labels, and placement within conceptual space. This was a period in which the laptop-like *Tablet PC* product was being marketed by Microsoft and their partners. One such variant of this product was the *Convertible Tablet*.

To further draw out the salient network features, I filtered the visualisation further by eliminating nodes with a degree value less than 19. This value is somewhat arbitrary, but is derived by using the average degree plus one standard deviation. By performing this further filtering operation, I am focusing on the nodes that are substantially embedded in the discourse. This is akin to highlighting frequent nodes, as happens with classic content analysis (Krippendorff, 2003).

Concept-network
 Tablet Sample
 2008-Q4
 Filtered, degree > 19
 98 nodes
 1200 edges

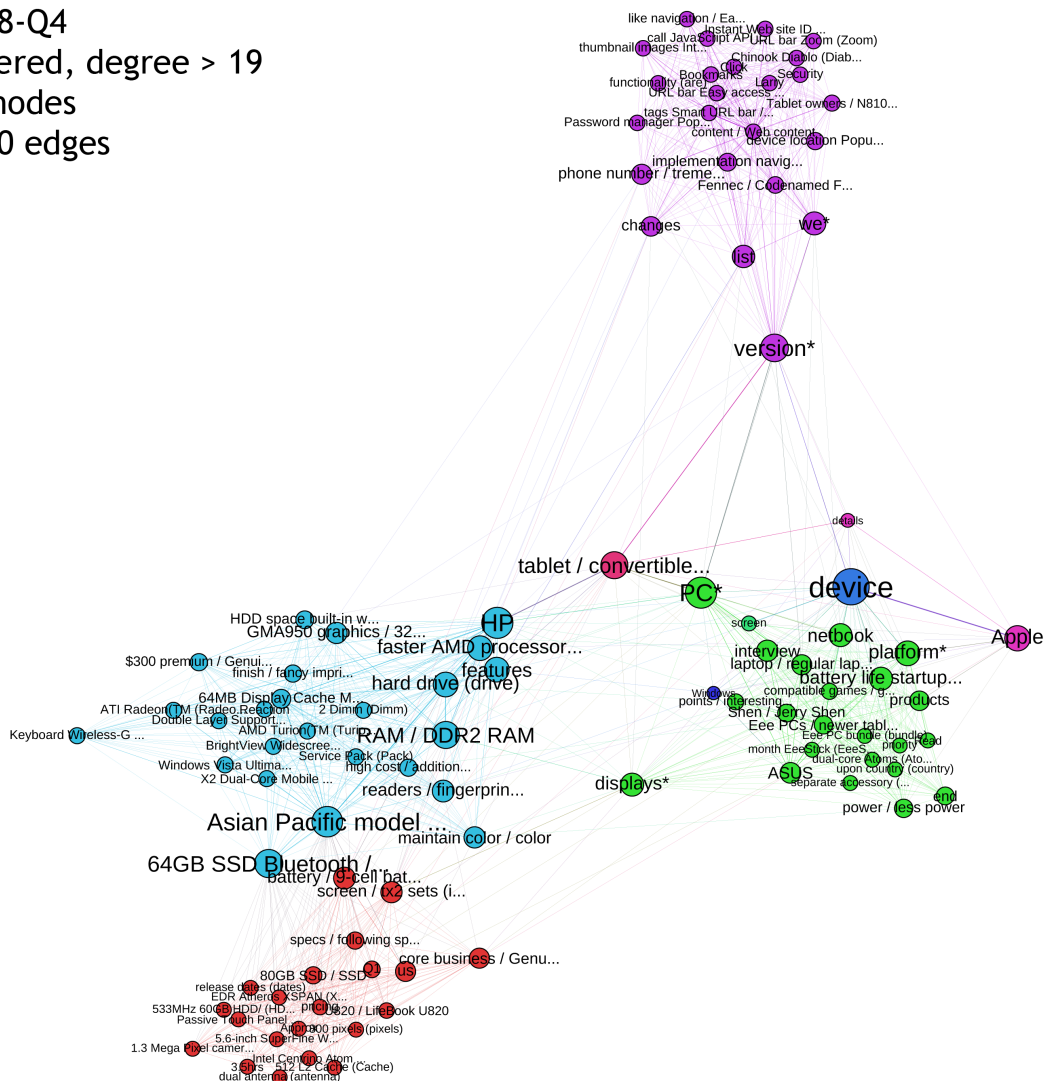


Figure 5.3: Concept network from “Tablet” Sample, 2008-Q4, filtered nodes, degree >

19

The filtered network visualisation for 2008-Q4 is shown in Figure 5.3. This network contains 98 nodes and 1200 edges. Here, the major clusters are far more pronounced than Figure 5.2. Along with the previous 4 clusters discussed, there is an additional purple cluster that appears to be connected to “PC” and “tablet /

convertible” through the node “version.” A cursory observation of this cluster shows nodes such as “like navigation”, “bookmarks”, “click”, “Password manager”, “content”, and “functionality.” In contrast to the red and blue clusters which are showing product attribute nodes prepended by numerical values, this appears to show concepts around functionality of a product.

These two visualisations afford us a rich description of a text network in conceptual space. Features of note are the clear clusters, their placement relative to one another, and the prominent nodes. Within the historical content, variations of Microsoft’s “Tablet PC” product are commercially available and the Netbook concept is just getting introduced. Note as well, that there is an Apple cluster, despite the company having never announced a tablet product. The Apple and Netbook clusters are positioned around the “device” node.

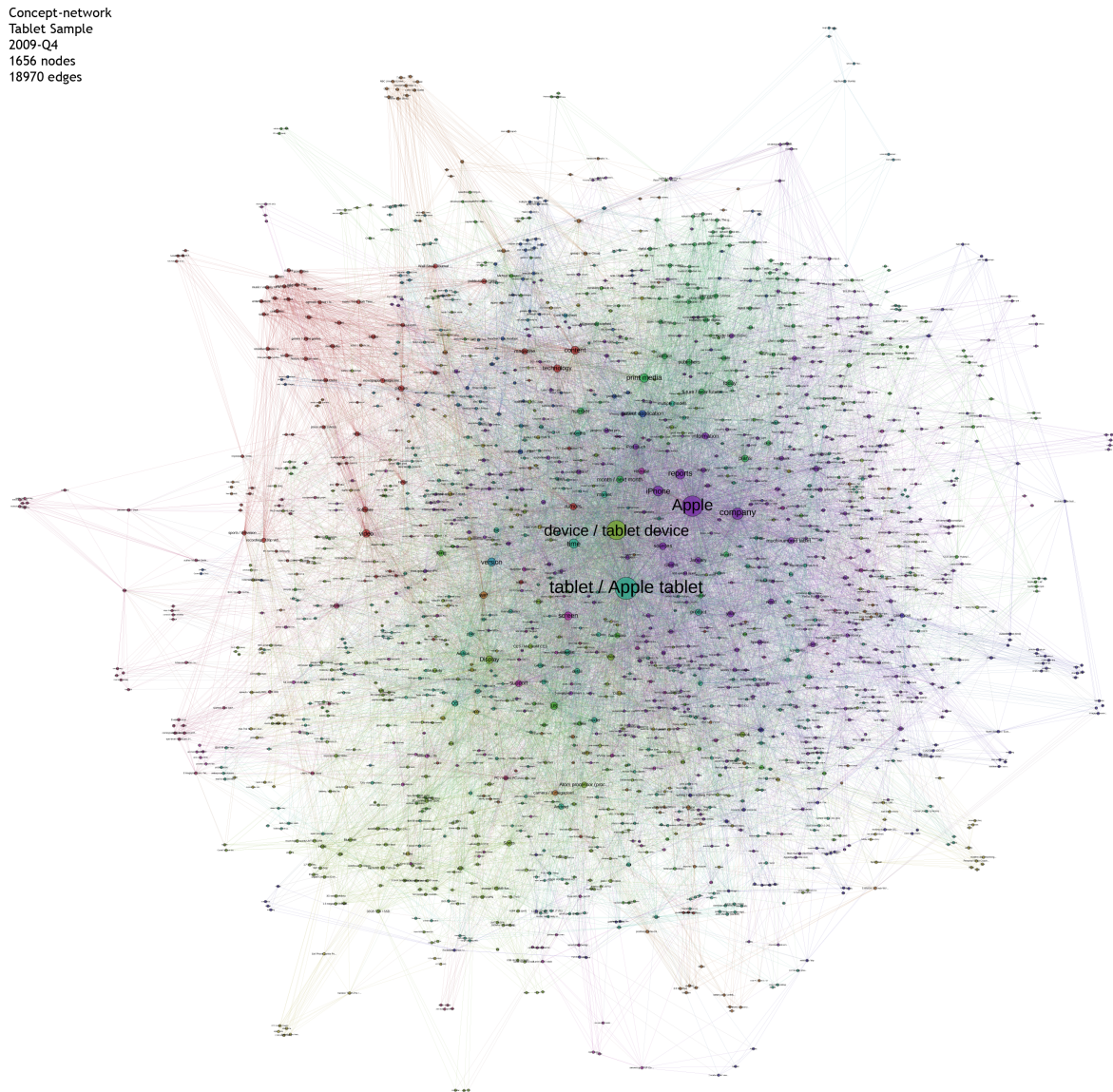


Figure 5.4: Concept network from “Tablet” Sample, 2009-Q4

Figure 5.4 reports on the concept network projection for 2009-Q4. In comparison to 2008-Q4 in Figure 5.2, this network is significantly larger and more dense. There are 1656 nodes with 18970 edges. There are roughly 4 partitions with much overlap between them. At the core, the most prominent nodes are “tablet/Apple tablet”, “device / tablet device” and “Apple.”

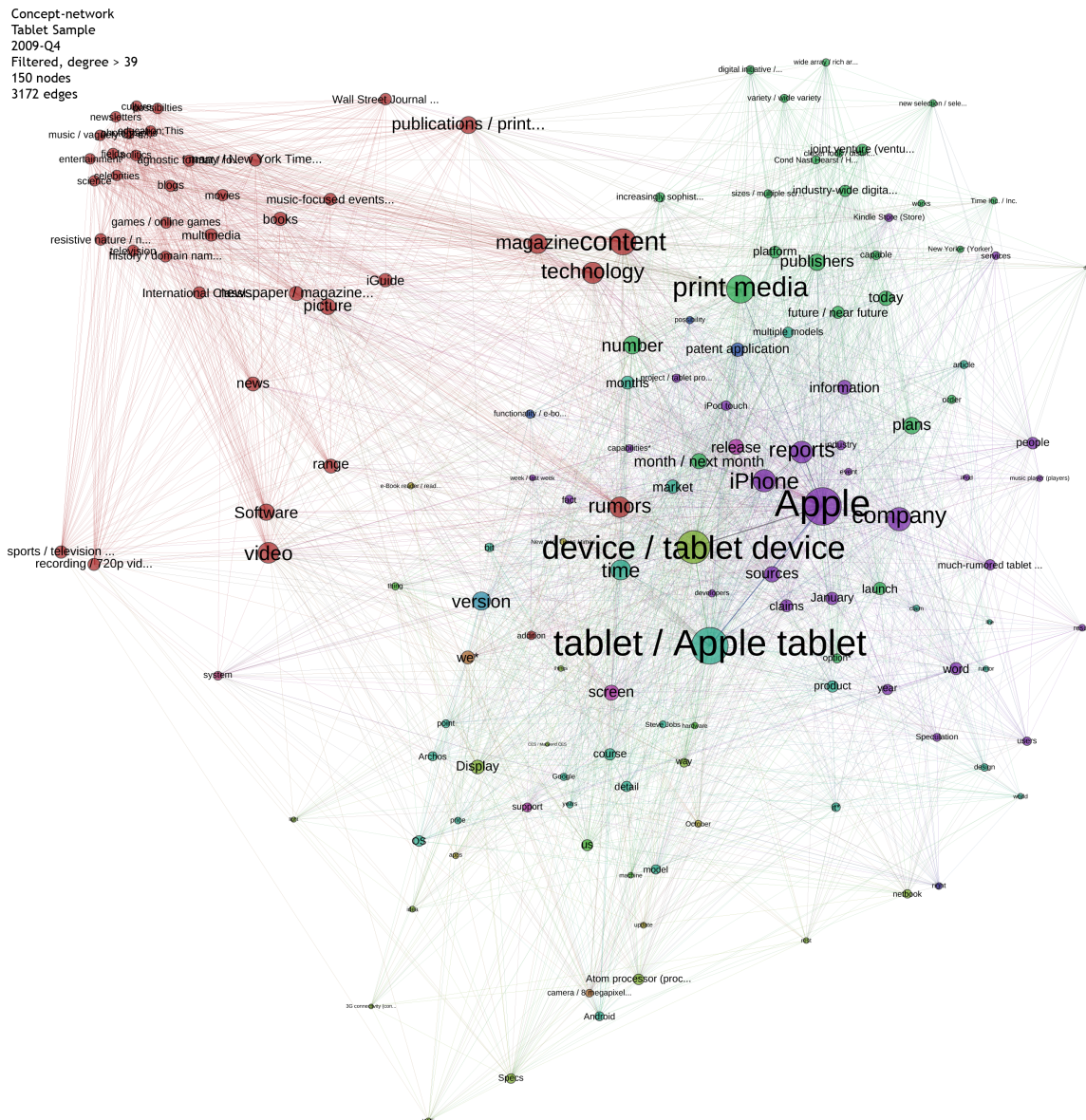


Figure 5.5 - Concept network from “Tablet” Sample, 2009-Q4, filtered based on degree > 39

In Figure 5.5, the concept network for 2009-Q4 is filtered down to nodes with more than 39 connections. Doing this allows us to focus on three key aspects. First, despite the filtering, there is significant overlap across the partitions. Secondly, there is a clear theme of an upcoming Apple tablet at the core of this network. The core of the largest partition has the following prominent nodes: “Apple”, “device/tablet

device”, “tablet/Apple tablet”, “rumors”, “iPhone”, “reports”, “release”, “next month”, “sources” and “launch”. Finally, the nodes central to filtered network projection tend to contain language which is general and ambiguous.

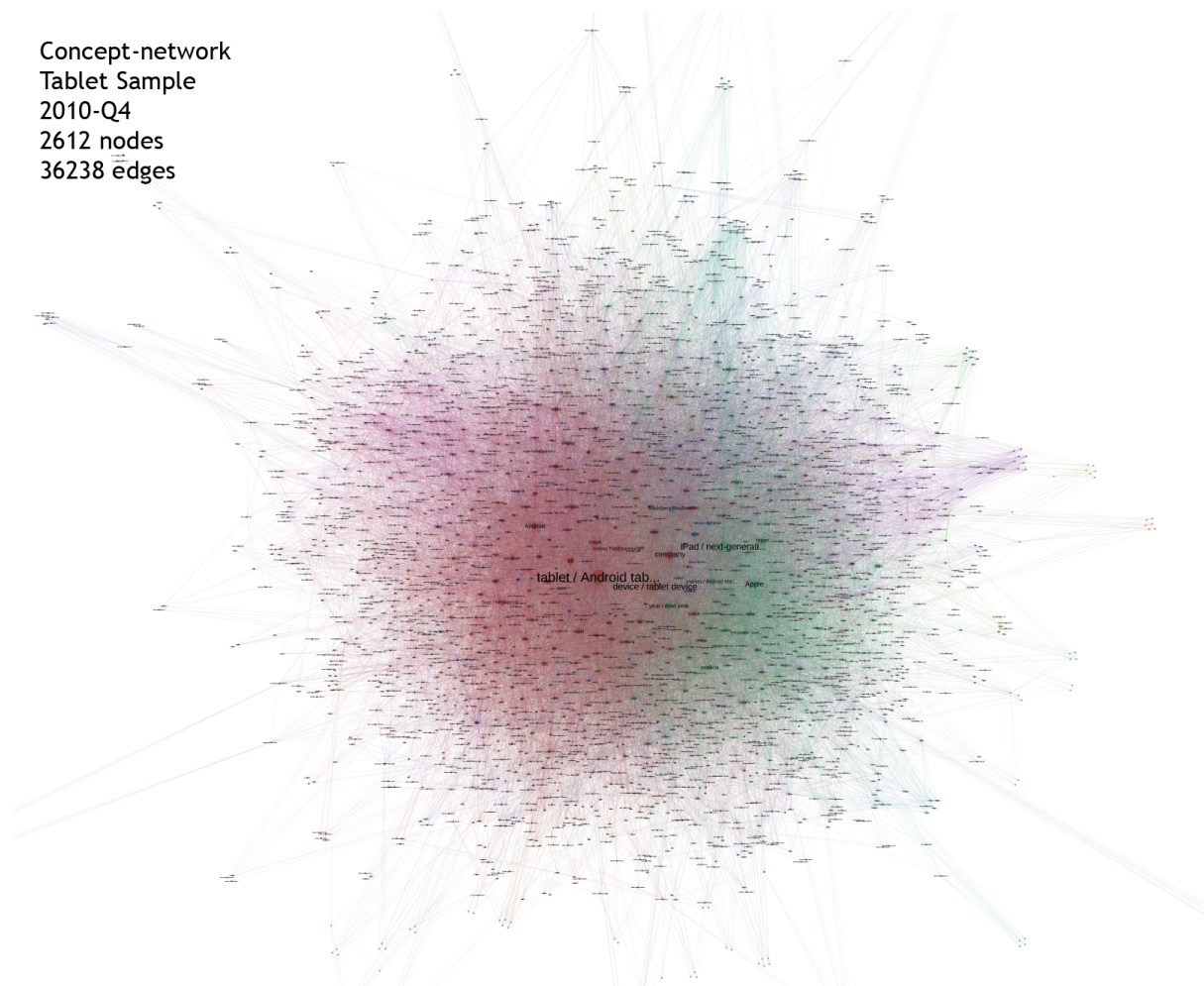


Figure 5.6 - Concept network from “Tablet” Sample, 2010-Q4

In Figure 5.6, the concept network for 2010-Q4 demonstrates discourses around the tablet after several major products were released. Notably, the Apple iPad, Samsung Galaxy Tab, HP Slate, and RIM Playbook had been publicly announced. As a concept network, this is dense with much overlap. Using a community detection algorithm, there are two major clusters that correspond to Android-based concepts

(in red) and iOS based concepts (in green). However, modularity is reported at 0.22, which is low. This indicates a level of stability in the discourse.

To get a clear picture of what the core concept associations are in 2010-Q4, I applied a degree filter of 96. Figure 5.7 shows the visualisation for concept network with this filter applied, showing a relatively sparse network of 137 nodes and 3891 edges. In this Figure, the two clusters become more apparent, but also another cluster emerges. In the top section of the visualisation, there is a cluster of nodes that correspond to product attributes. This has been blown up in the inset. Interestingly, these nodes are associated and straddle the major modules of Android (Google) and iOS (iPad).

Concept-network
Tablet Sample
2010-Q4
Filtered, degree > 96
137 nodes
3891 edges

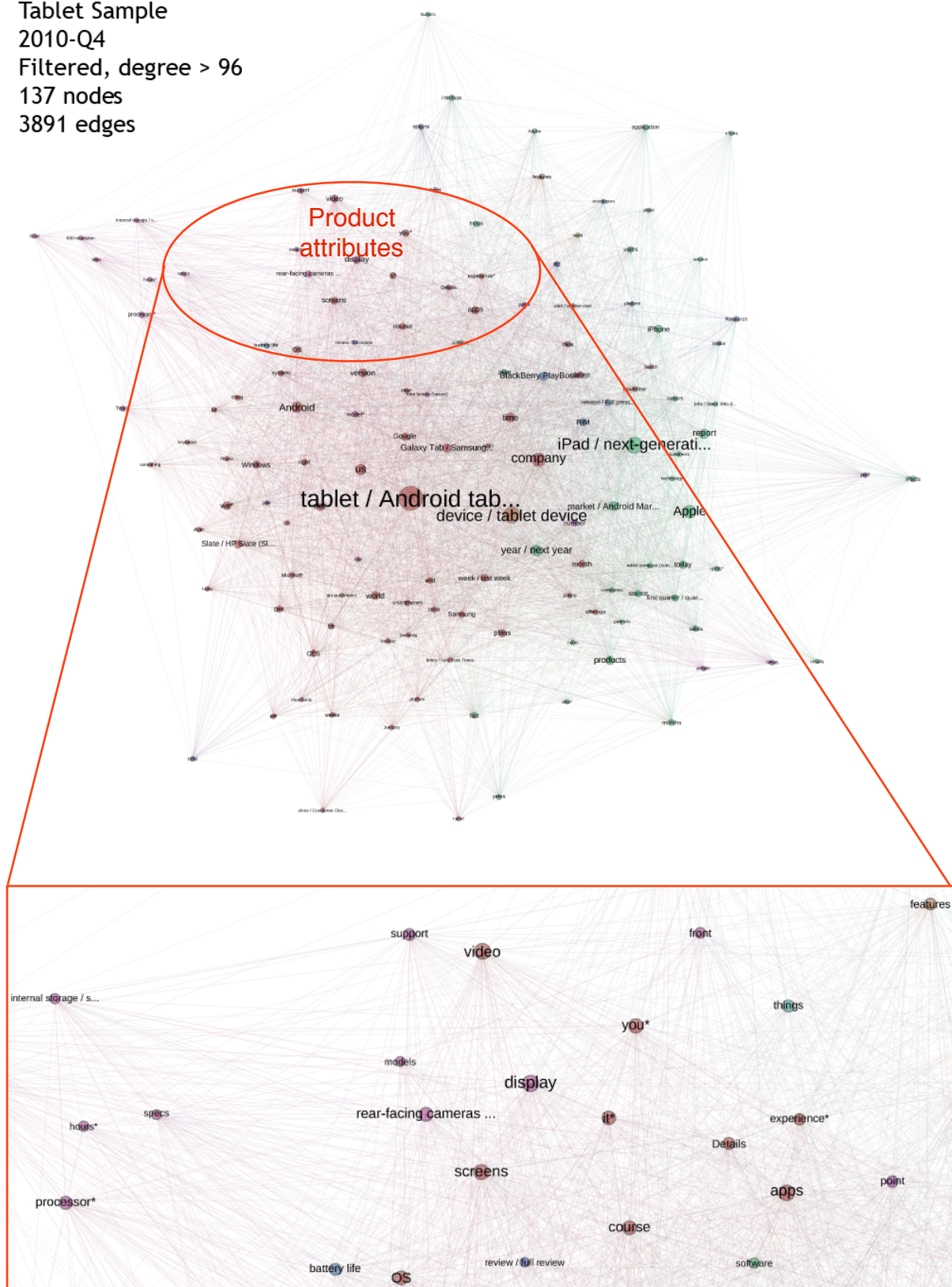


Figure 5.7: Concept network from “Tablet” Sample, 2010-Q4, filtering based on degree

> 96

To characterise the major clusters, I conducted a further degree filtering, setting the threshold at 217. Figure 5.8 reports the visualisation for the filtered concept network. Here, descriptively we see “device / tablet device” as the most central node. Android tablets and Apple iPad feature predominantly here. However, we also see other products, “Blackberry Playbook”, “Samsung Galaxy Tab”, “HP Slate.” On the top left side of the network is a cluster of product attributes. The node “tablet/Android tab” exhibits the highest betweenness centrality. This indicates the extent to which this links the flow of other concepts.

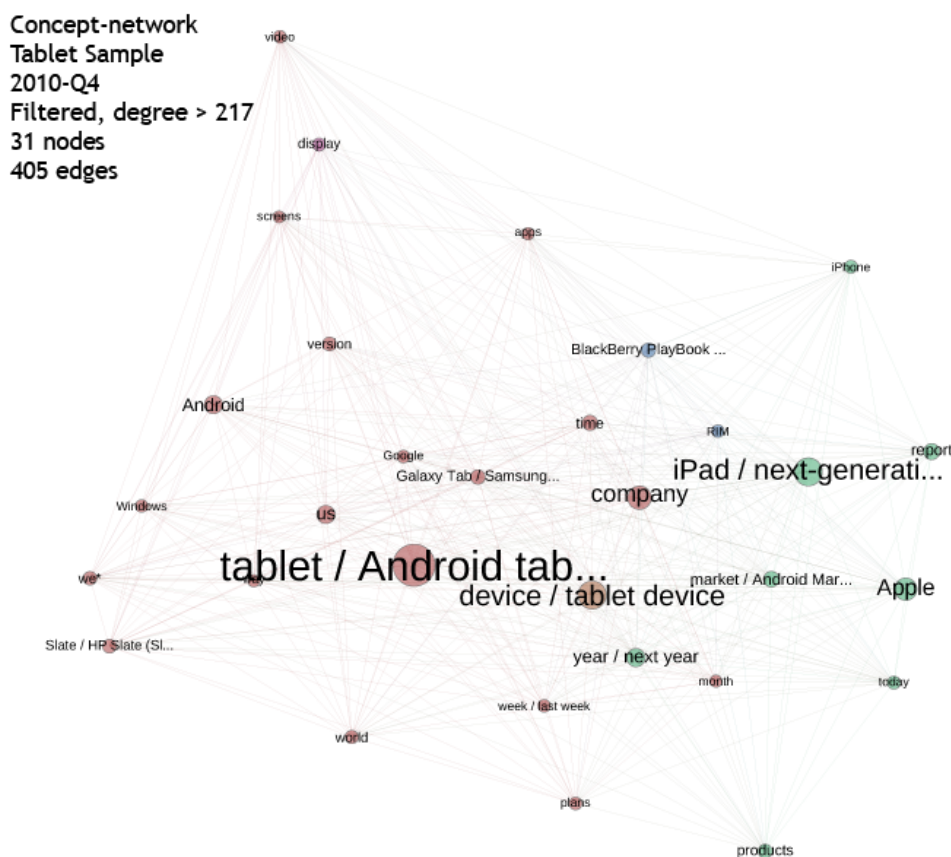


Figure 5.8: Concept network from “Tablet” Sample, 2010-Q4, filtering based on degree > 217

The three concept networks were analysed primarily for clusters and prominent nodes indicating themes in a relational conceptual space. This is useful for developing a grounded understanding of themes that connect across the corpus. Early on, there seemed to be a “tablet PC” frame that was dominating the discourse, then a year later “Apple tablet” seemed to be central.

What these analysis do not show is a systematic analysis of the structure of the networks. As in social network analysis, analysis of these networks occurs as two levels, the first is individual nodes and their relational characteristics. The second is inferences regarding the structure of the network as a whole.

Recall from Chapter 3 how the network measure modularity was used to show cohesive a network structure was. I argued that when this measure was applied to discourse networks, this can measure how connected the discourse is. To make inferences about an emerging product ontology, I apply a modularity measure to the quarterly concept networks taken in this sample.

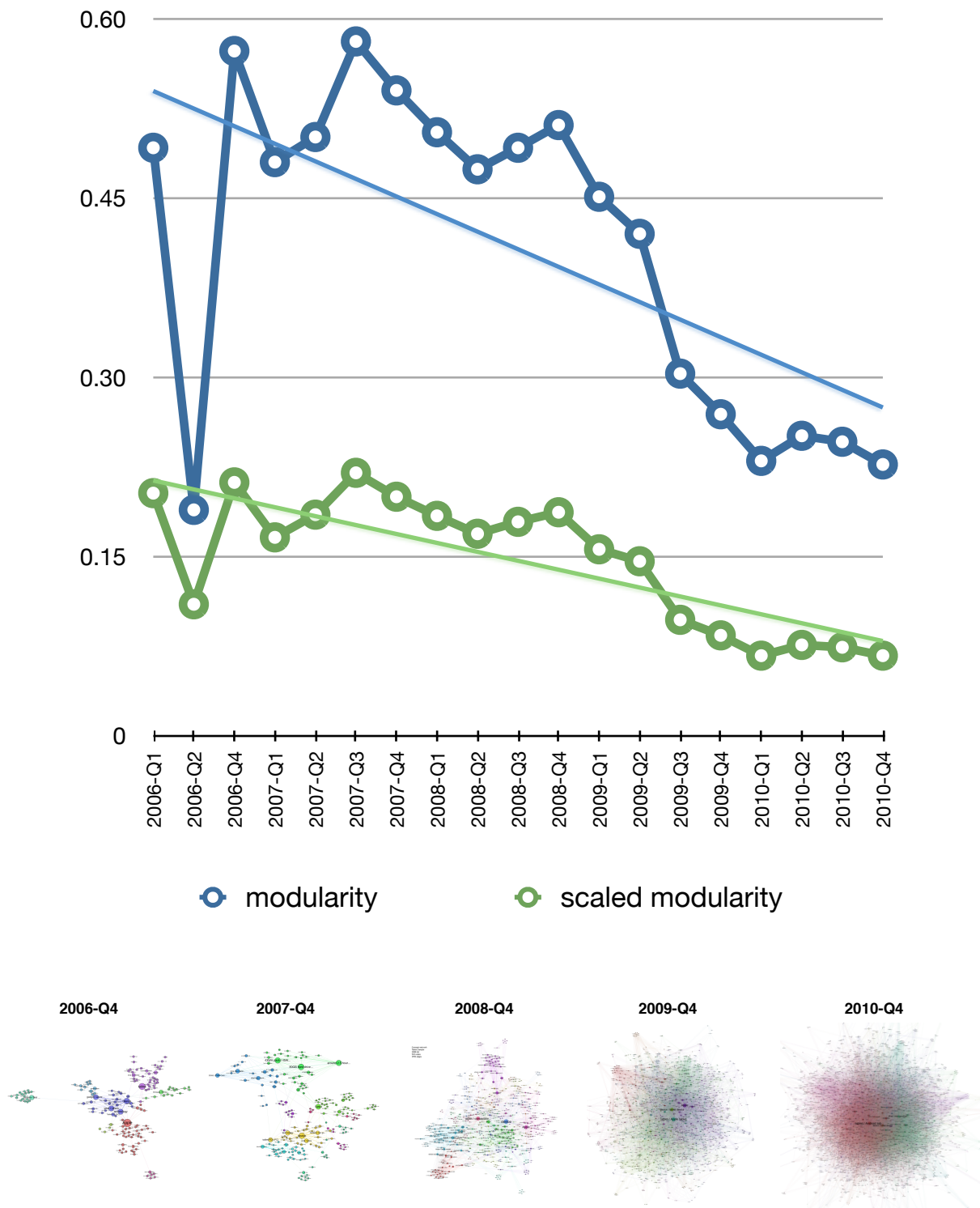


Figure 5.9: Modularity Scores For Overall Concept network from “Tablet” Sample & Network evolution

Figure 5.9 reports the modularity scores for quarterly concept networks along with trends. Raw network modularity was used here, but also a scaled modularity measure. The modularity algorithm was presented as not containing any scale biases (Newman & Leitch, 2006), however Shwed & Bearman (2010) cited work done on biology networks that found that modularity increased with network size (Kreimer et al., 2008). To correct for this possibility, Shwed & Bearman (2010) divided the modularity by the log of network size to get a scaled modularity measure. Both are reported in this figure.

In Figure 5.9, there is a trend of an overall decline in modularity scores over the quarterly networks. This result is the same for both modularity and scaled modularity. There is some fluctuation throughout the same, but from 2008-Q4 until 2010-Q4 there is a noticeable drop. In particular, in between 2009-Q2 and 2009-Q3 sees the largest drop in modularity throughout the sample.

The major Innovation Disclosure Events (IDE) by Apple and Microsoft occurred in 2010-Q1, where the iPad and HP Slate tablets were introduced. It is understandable that volume of market stories would pick up after these events (Rosa et al., 1999). However, extant theory does not adequately explain how this conceptual stabilisation could occur beforehand.

In an exploratory study like this, an interpretation of a drop in modularity scores of concept networks is not trivial, however Chapter 3 demonstrated that lower modularity corresponded to a more connected, cohesive network. A higher

modularity network would signify more clearly partitioned clusters. As more broker nodes are added, there is more overlap and thus a lower modularity score.

For a relational view of a product ontology, this result is compelling evidence of a stabilising nomenclature. Furthermore, the data suggests that there are two phases with a transition occurring between Q2 and Q3 2009. In Figure 5.9, we see this is the largest drop in the sample. Going back to Figure 5.1, the same period appears as a breakout of posts containing the word “tablet.” Notably, this period is at least two quarters before the major public announcements from Apple, Microsoft, Google, Samsung and Research in Motion regarding tablet products. This is counter to the literature, which emphasises that Product Ontologies are initiated by producers who act as “ontological critics” and release different designs of products into the market that then initiate market stories. This empirical evidence leads to my first proposition about the phenomenon,

Proposition 1: *A nascent product ontology emerges and coheres within a product public prior to Innovation Disclosure Events (IDE) by producers.*

5.3.2 Coding for cognitive classes

The methods section in Chapter 3 outlined how a relational view of an empirical product ontology projects unit level concepts in network form. Each concept has centrality measures in the network that indicate their placement. I focus on betweenness centrality (Freeman, 1977), drawing from Centering theory (Grosz et al., 1995) in linguistics to demonstrate cognitive embeddedness (Zukin & DiMaggio, 1990). I coded the top concepts according to their cognitive class within the product ontology. This enabled me to trace the broad elements of the cognitive structure as it emerges.

For each quarter in the sample, I coded the top 30 concepts and gathered sums of centrality scores for each cognitive class. Figure 5.10 reports on these values per quarter after normalising based on the total sum of all centrality scores for the period. This chart maps the cognitive classes as aggregates and demonstrates the level of focus for each.

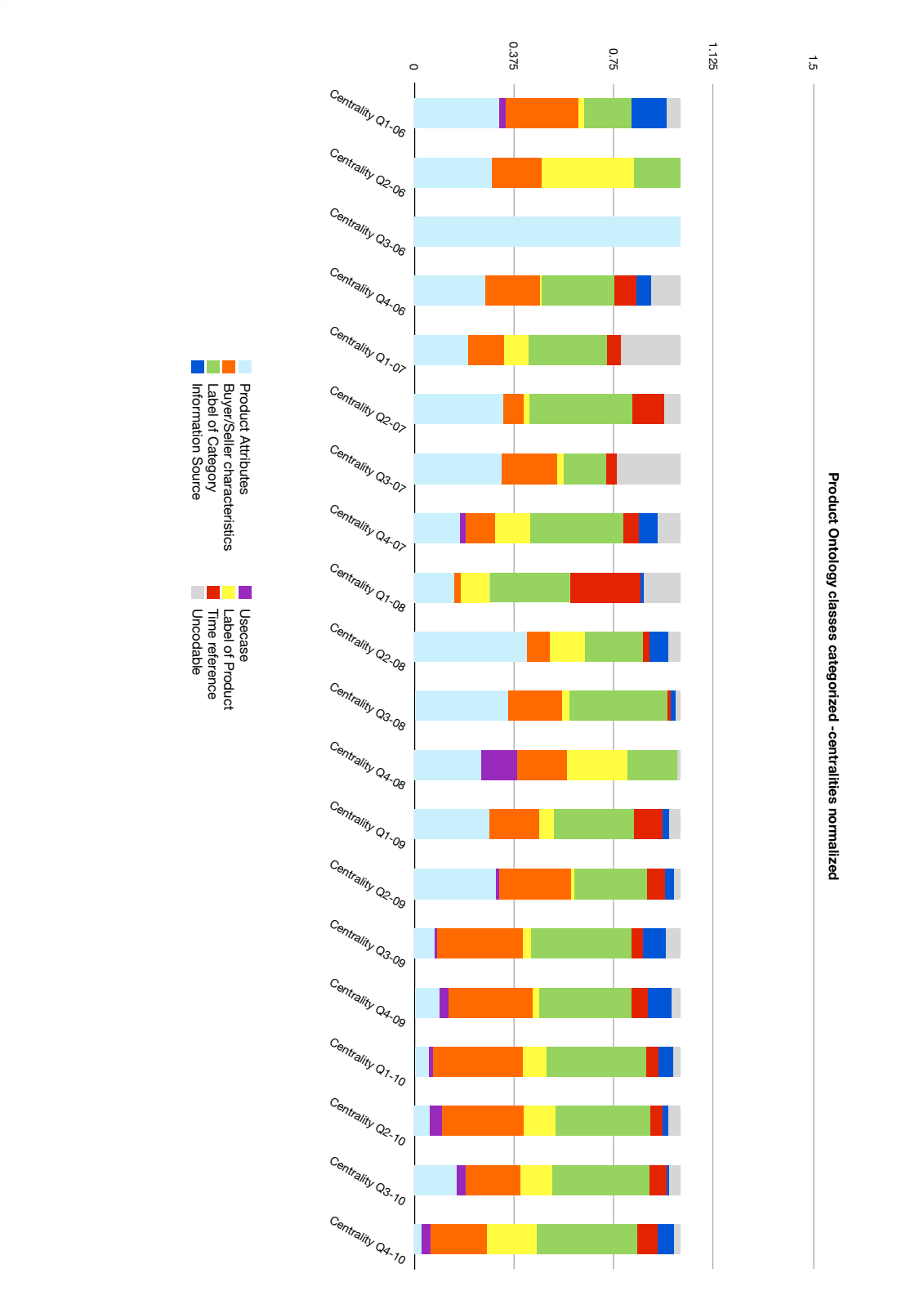


Figure 5.10: Distribution of Cognitive classes in a product ontology

Figure 5.10 allows us to see three major things. First, the class *Product Attributes* drops over time. Second, the class *Buyer/Seller characteristics* rises over the sample. Finally, we are able to infer stability occurring in 2009-Q3. I explain each below in more detail.

Product attributes are coded here as the specifications of a product. As seen prominently in Figure 5.3, they refer to the elements such as *RAM*, *display*, or *Intel Centrino Atom Processor*. As a cognitive grouping, this appears to be central and thus important in early discourse of the sample. In Figure 5.10, this appears initially central and drops in 2009.

Buyer/Seller characteristics are coded here as nodes that refer to social actors in the market. The analog of this is the Kennedy (2008) work on the cognitive embeddedness of producers within media stories. However, here I group buyers and sellers together to distinguish social from other cognitive forms.

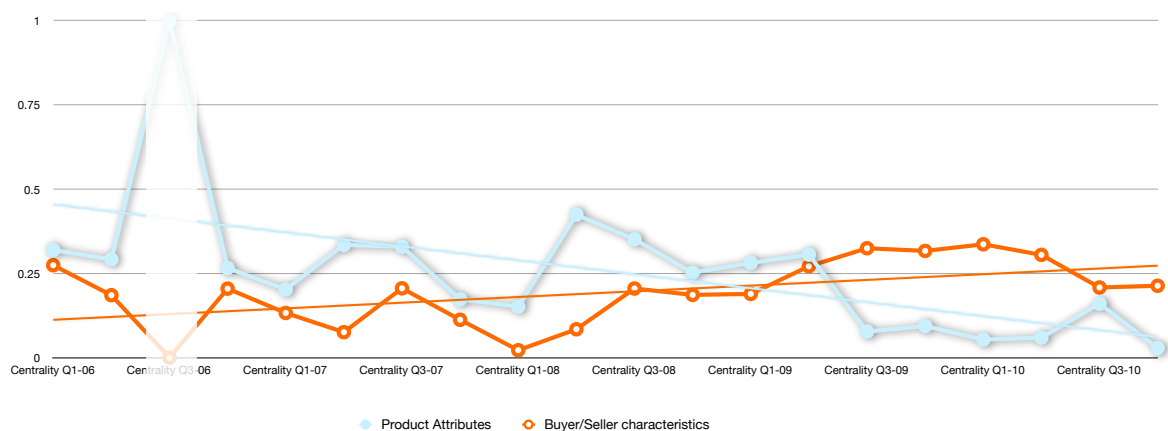


Figure 5.11: Distribution of Cognitive classes in a product ontology - trends

In Figure 5.11, the *Product Attribute* and *Buyer/seller characteristics* classes were plotted alongside trend lines. Over the course of the sample, the product attributes drop in centrality, whilst the buyer/seller characteristics rose. The transition point between the two occurs in the interval Q2 -> Q3 2009. This is a notable date as it matches the largest drop in modularity in Figure 5.9 and is evidence of a move towards stability.

To find further evidence of stability leading to a product ontology, I attempted to find when these cognitive orderings stabilised. One way to show this statistically is run a chi-square test for independence. I ran this test for each quarter and the following one. The results from these test showed the only period in this sample that not independent was Q3 -> Q4 2009. This is evident visually as well as quarters ranging 2009-Q3 -> 2010-Q4 resemble similar a shape.

With the IDEs by Apple and Microsoft in 2010-Q1, we see a matching rise in the number of *tablet* articles in Figure 5.1. However, the evidence of stability in the concept network points to 2009-Q2, two quarters before the firm's IDEs. This evidence leads to the following proposition,

Proposition 2: *The cognitive classes in a product ontology stabilise within a product public prior to innovation disclosure events (IDE) by producers*

In Figure 5.11, across the sample there is a trend of *Product Attributes* dropping and *Buyer/Seller characteristics* rising. These two measures finally cross in 2009-Q2.

This date corresponds with the stabilisation from both the network modularity score and the cognitive class orderings. This leads to the following proposition,

Proposition 3: *Buyer/seller characteristics will replace Product Attributes as primary importance when a nascent product ontology stabilises*

This empirical result of the concept network around *tablet* stabilising prior to the firm IDEs and volume increase in market stories is an important finding in tracing the emergence of a product ontology. Counter to extant theory (Rosa et al., 1999) that suggests the product ontology stabilises in line with a volume of market stories increasing, there is evidence here to suggest that this stabilisation is occurring at the level of the concept network and cognitive class ordering. This leads to the following proposition,

Proposition 4: *There is stability within the cognitive classes of a product ontology even during increased attention to a nascent product ontology*

Together, these propositions help to build a picture of a product ontology emerging within a discourse collected across four rather separate sites who settled upon a sufficiently cohesive discourse. This represents a *product public* deliberating based on a discourse that is firmly within the realm of speculation and rumour. Although early on in the sample there were alternative products such as *Tablet PC* being marketed, the *device* and *Apple* frames were clearly emerging. The shapes of these concept networks speak to the meaning structure of the *tablet* product ontology and this was seen within a *product public* before the major firms put on IDEs.

Chapter 6: Discussion & Conclusion

6.1 Introduction

This chapter discusses my contributions to a discursive view of product market emergence. In this dissertation, I developed propositions around a *product ontology emerging in a nascent market*. For the empirical work, I used a novel method of concept networks to find meaning structures from tech blogs. I learned that front page activity on a tech blog showed structural signatures consistent with the definition of product publics. This has implications for the dominant perspective of community within innovation and organisational studies. I also learned that meaning could be measured in a concept network, and network modularity was good proxy for measuring its cohesion in a text corpus. This enabled empirical work on rumour and speculation about the tablet for 6 years leading up to the Innovation Disclosure Events. The implications for innovation theory will be discussed, along with the limitations.

Extant theory explains how the product ontology develops in market stories after the inventing organisation serves as an ontological critic and aims to solve design problems. This is in line with cognitive approaches to the technology lifecycle literature (Tripsas, 2008) that consider process after producers have released products into a developing market. In this dissertation, I set out to explore the proposition that *product ontologies begin to develop before commercial release through product publics*. I addressed this through an empirical study of tech blogs discussing the emerging tablet market.

This approach to theorising product ontology emergence is based on a phenomenon that has surfaced in recent years with social media. Drawing upon previous ideas of user innovation and online community literatures, the *product public* is a collectivity that is temporarily constituted with discourse. It is neither a community or a crowd, but a third category. I reached back into pragmatist sociology to develop a specific take on a public. This deserves to be studied because, on the one hand, it is similar to communities and crowds; but it is also quite distinct. Its distinctiveness is what makes it important. As a social entity it is not static, but rather fluid as resources flow in and out over time (Faraj et al., 2011). When analysed at the aggregate level, these dynamics do demonstrate some structural signatures of activity, but particularly so when the content itself is analysed. Its importance is in how it helps to set expectations as an arbiter of product ontologies. This thesis engaged in both the process of delineating the nature and definition of the *product public* and its characteristics, through a new computational approach to discourse analysis.

To address the research question, I first had to find where and how the *product public* was constituted. To do this, I began by thinking about tech blogs as potential sites engaging in loosely bound communication, due to the ease of entry, the relative anonymity, and domain-specific attributes of the high-technology space. In Chapter 4, I undertook an in-depth analysis of a full discussion forum over a 13 month sample. I looked at discussion dynamics by users and found evidence of a *product public* as an organising principle that was different from a community. The results showed that a *product public* seemed to be constituted by an audience forming around front-page articles from tech blogs. Thus, the articles were taken to be a

proxy for the active members of a *product public*. I then endeavoured to trace the emergence of a product ontology through a *product public*. To study a product ontology in discourse, I operationalised it through a relational perspective as a meaning structure. This required a novel methodology for studying the structural aspects in network form. Chapter 3 dealt with developing this method in detail, which enabled me to undertake work in Chapter 5 to analyse the emergence of a product ontology. The results from this chapter showed evidence of the tablet product ontology meaning structure emerging in tech blog discourse. Stabilisation of the meaning structure was found through modularity measures and the centrality of cognitive class orderings. Interestingly, this stabilisation occurred several annual quarters before companies officially announced their intent to release a product in that space.

6.2 Implications and Contributions

The critical point I develop in this research is a pragmatist view of nascent markets. In a sequence of market development, both the "sociology of markets" and technology lifecycle explanations apply as downstream mechanisms. I theorise that upstream there is meaning construction work done through cognitive embeddedness of social actors as well as forms. This extends our idea of cognitive referents to include associative forms (Santos & Eisenhardt, 2009). Within the Economic Sociology literature, this is akin to expectations of demand beginning to stabilise before the market properly emerges. These can be seen as "collective technology frames" (Kaplan & Tripsas, 2008), or expectations that will later provide the basis for a shared knowledge structure between those who will eventually make up the market network when the field settles. Therefore, this meaning construction is the product ontology developing before market stories, and before social networks of an organisational field. In a nascent market, as a pre-institutional arena, market information is not backed up by credentials that afford legitimacy. The effect of lessened legitimacy in a pre-institutional arena leads to rumor being deliberated in discursive open contexts as uncertain knowledge. This is similar to Shibutani's (1966) model of rumour as a collective sensemaking transaction.

I showed that rumour and speculation in a nascent market do begin to form into cognitive structures. From this, I theorise that innovation concepts are linked in a discursive web that may reach temporary stability. Innovation concepts have been studied as collocations (Wang, 2009), and in this thesis, I extend them to the conception of a *relational product ontology*. This provides an important contribution

to the literature on new structuralism that is attempting to link stories and narratives with the formation of a field (Lounsbury and Ventresca, 2003; Mohr, 1998).

A second argument that I developed is that in the evolution of a product market, before "sociology of markets" and economic explanations apply, there is cognitive work done in the collectivity of a *product public*. This is not necessarily made up of strong ties or shared identity and, thus, is not a community. Wang & Ramiller (2009) argue that innovation-concepts are defined in discourse known as an "organising vision" by a community of social actors. Legitimacy and community are linked concepts. This was explored in Chapter 4, through a detailed empirical investigation of dynamics in an online forum. There is evidence of a community present, but it is not a sufficient construct to explain the dynamics in that setting. Similar to the argument Anand & Peterson (2000) develop in the concept of a market information regime, the *product public* is an arena that serves as a loose "infomediary" (Deephouse & Heugens, 2008). It is a discursive arena, similar to a social movement or discursive-fields (Hoffman, 1999; Snow, 2004). As Tarde and Dewey describe, the public is made possible through mass communication and is effectively an arena for the development of propositions as uncertain knowledge (rumours) about an emerging product market.

6.2.1 Product Ontology

One theoretical contribution of this work is towards a further understanding of the product ontology. The main research question I developed in this thesis was, *how does ambiguity get resolved in the early moments of nascent markets?* When the discursive view of market emergence was applied, this translated into the sub research question, *how will a meaning structure stabilise relative to a field configuring event?* The product ontology was proposed as one type of meaning structure. Recall that Porac et al., (2002), refer to the product ontology as cognitive representations packaging up product attributes, use-cases and buyer characteristics. These knowledge structures are shared by buyers and sellers as a joint understanding that evolves over time. This built from the Rosa et al., (1999) study of schemas containing a constellation of attributes that made up the minivan, and relates to a literature that is only beginning to explore how new schemas can develop at the collective level and be sustained through cultural categories (Bingham & Kahl, 2012; Dougherty, 1990). Zerubavel (1999) asks why we are able to combine cheese and a hamburger to form a cheeseburger, but the same cannot be said for a “ketchup-burger.” He emphasises how our social environments deeply affect how we perceive, classify and assign meaning to things. As ‘thought communities’, they mark out a structure that enables us to perceive things and classify reality. The process of classification entails the creation of symbolic associations that involves shared meanings (Zerubavel, 1999: 69). This relates to a relational product ontology as a theory of meaning around markets as cognitive structures. Faraj et al., (2011) describe a loose definition of online communities whereby resources flow in and out and there may not be a similar set of ontological assumptions. This helps to extend

our thinking towards *publics*, but I demonstrate here that perhaps publics can cohere around a meaning structure that resembles a relatively stable ontology.

One proposition developed in this thesis is that the product ontology could be considered from a relational view. This operationalises the ontology as a web of meaning that draws together an understanding through the notion of *family resemblances*. From the works of Ruef (1999) and Wittgenstein (1958), I describe the product ontology as a fuzzy set of similarities and differences. Unlike the relatively static model of dominant design (Tushman & Anderson, 1986) that suggests that markets achieve near-perfect consensus with strict boundaries, this model is able to account for the field as a cognitive web of overlapping similarities. This is a wider perspective of a market as beyond simply social relations between producers. It brings together a local cultural view of product markets as a series of shared belief systems that become embedded over time in a layered structure.

Recall from Chapter 5 that the tablet product attributes and producer network shifted their centralities over the course of the emergence process. This is particularly important in addressing the messiness of overlapping product ontologies. The tablet case is a good example of an old product ontology overlapping a new one. The tablet PC as defined by Microsoft in 2002 was still present in discourse until 2008. The ontology of the “post-PC” tablet that we know today by the iPad or the GalaxyTab, is neither homogenous or cleanly bounded. However, it does exist in the minds of producers, consumers and other market actors. Similar to Bingham & Kahl’s (2012) study of the computer schema emerging in discourse of the insurance industry, there is a loose cognitive structure of associations that links the tablet

products into a similar space. The product models may not be identical, but they do share a common nomenclature.

The claim of shared meanings' central role in sense-making is linked to the idea of culturally available schemata, as

“knowledge structures that represent objects or events and provide default assumptions about their characteristics, relationships, and entailments under conditions of incomplete information” (DiMaggio, 1997: 269).

These schemata draw from the general cultural toolbox, and in turn become encoded within the local or broader culture. These shared schemas may then become the substance of what Marshall (1890) and others (Saxenian, 1994; Shibutani, 1966) refer to as being “in the air.” This links to work being done on a cultural approach to entrepreneurship (Lounsbury & Glynn, 2001). It is at this point that legitimacy of the ecology of innovation concepts plays an important role in establishing forms of meaning structures. New products are released to a public that has already established schemas within a nascent product ontology. This challenges the evolutionary economic view of demand in an early market as being inchoate (Geroski, 2003).

Another proposition developed here is that innovation rumours may initiate product ontologies whilst inventors and inventing organisations are still attempting to form them. Extant theory suggests that product ontologies as meaning structures begin in market stories that are told after producers release products and buyers attempt to make sense of them. While this still may be true, I have shown evidence to support the argument that the initiation of product ontologies is occurring in a market

information regime made up of rumour and speculation. A *product public* is made possible because of tech blogs and social media. Firms and other eventual social actors in a field are able to follow or engage with a *product public* without having strong social ties because the barriers to information are so low online. As a collectivity, in a *product public* it is not facts or boundaries that are transmitted, so much as the discourse that may begin to structure a nascent product ontology.

Thinking about an empirical product ontology is only powerful insofar as a sufficient level of stability emerges. Amidst the messy ecosystem of innovation concepts, webs of meaning get strung together and can serve as a nomenclature that enables sensemaking. When approaching an innovation anew, social actors attempt to place it within their existing set of available schemas (Clark, 1985). This analogical reasoning allows them to see the new object against the understood objects, drawing upon the raw material and cues that afford understanding and sensemaking (Weick, 2005). Research on social cognition in a market has attempted to derive a complementary explanation of innovation sources being that of shifting customer preferences, industry frames, and categories (Porac et al., 2002). Tripsas (2008) develops the concept of customer preference trajectories as being in parallel with technological ones. In a longitudinal study of the typesetter industry over a centurylong period, she demonstrates how market demand evolves and pulls innovation in a particular direction. However, beyond Bingham & Kahl (2012), very little theorising has been done on the emergence of such cognitive structures.

Tripsas (2008) outlined that shifts in customer preferences are shown to be due to shifts in the socio-political environment, shifting interdependence of a modular

system, evolution of customers over time through increased size and complexity, and proactive moves by producers to shape preferences through the media. Kennedy (2008) also investigated how media affected legitimacy for firms in an emerging industry by signalling embeddedness. Understanding the network of competitors meant drawing a boundary around a constellation of firms. Together, this constellation marks a cognitive boundary around an industry. However, this focus is still largely on social relations as the skeletal structure. In this thesis, I described how discourses can take on structures as well and these can include meaningful associations beyond social relations. A *product public* is offered as a collectivity that creates, contributes to, and is imagined in discourses.

6.2.2 Product Publics

It is notable that the empirical data in this thesis centred around Apple and its iPad tablet. Not only is Apple one of the most valuable companies in the world, but it is notorious for attracting a fanatical following (Kahney, 2006). The brand community (Muniz & O'Guinn, 2001) has been used as a construct to make sense of behaviours of the Apple faithful (Muniz & Schau, 2005). According to this perspective, members are supposed to have tribal allegiances and celebrate retelling of the brand narrative (Muniz & O'Guinn, 1995). I attempted to explore this dynamic in Chapter 4, by trying to identify users on the large web-forum *Macrumors.com*. Considering the brand community frame, I expected to find a node of *hard-core* users that heavily engaged in the same set of discussions around Apple rumours and deliberated them like an online public sphere (Habermas, 1992). What I found was evidence of some

community and support activity, but there also was large influx of discursive activity from an assortment of users who would only rapidly post in specific forum categories regarding news and perhaps rumour. This, I use to help us think about a public as being distinct from intensely linked fans with tribal connections to a brand. It is perhaps fitting as a critique of the brand community literature that this *product public* is appearing on the largest Apple forum on the Internet. Though it should be stressed that a public is not the same as *the mainstream public*, and as conceived of here, is distinct from the public sphere. I will discuss more in articulating the product public below.

By the early 20th century, publics were central to American democratic theory. Made up of “omnicompetent” citizens who possessed a similar mind and will, the public was depicted as an institutional organism that wielded much influence over morals and events (Lippman, 1925). As a Platonic ideal, the public was knowledgeable, influential, and able to guide societal affairs over the layperson. In a critique of this, Lippman (1925) wrote about publics as being Phantoms - that is, myths of idealism. He suggested that although publics existed to some extent, society was primarily made up of either agents or bystanders with the former able to act as “insiders”, and the latter merely there to watch as outside spectators. He reasoned that the public was not a rational group, and only really exerted power through the means to elect “insiders” into office.

Dewey (1927) wrote a response to Lippmann (1925) where he defended and elucidated the concept of a public as a deliberating group. He described how a public comes together, as individuals having been concerned with an issue and concerned

with the possibility of being affected by its consequences. This depiction of publics described three stages of a development. First, the somewhat separate individuals find that they themselves face a similar indeterminate situation. Second, they recognise what is indeterminate about the situation and define the problem. Finally, they organise to do something about the problem. Dewey's publics were about deliberation; as people came together concerned with an issue, they were also concerned with the possibility of being affected by consequences.

The group that came to be known as the second "Chicago School" of Sociology (Fine, 1995) took Dewey's publics further. Blumer (1948, 1966) distinguished between mass opinion and that of publics, reasoning that the latter were a homogenous social group that got confronted by an issue but initially have differing ideas on how to address it. The public engages in discussion and deliberation about the issue. In his study of rumour as a form of improvised news, Shibutani (1966) followed Allport & Postman's (1947: 180) assertion that, "every rumour has a public." He also followed Dewey (1927) and Blumer (1948, 1966) in suggesting that rumour publics are skeptical, rational and engage in some form of critical deliberation; that is they check assertions and attempt to verify the reliability of the sources. This type of rumour public forms in certain conditions and conduct a social transaction deliberating on plausible propositions. More recent work has begun to consider publics as temporary discursive collectives (Mische, 2009). White (1995:1053) defined a public as

"a linguistically constructed interactional space which is an approximation to a non-subjective, maximally decontextual, interactional setting exhibiting fully connected actors, whether considered equivalent or arrayed in transitive pecking order. "

This definition emphasises the discursive interactional notion of the public as a space.

6.2.3 Community

The concept of community has been used in many ways. As an English word with almost no negative connotation, it is often used as an aspiration for belonging and doing good (Cohen, 1985). In a review of sociology, Brint (2001) found historical record of what has been a deeply problematic analytical concept. As far back as Durkheim [1897] (1997) there overlapping ideas over community, social groups and society. The sociological categories of *Geimenschaft* and *Gesellschaft*, were created to describe the differences between communities with tight kinship, social mores and division of labour. In the *Geimenschaft*, community is seen to have tight kinship and strong social mores taking precedent over individual. *Gesellschaft*, on the other hand, was an association where the individual took precedent over the group (Brint, 2001).

By the 1950s, in sociology there were 94 separate definitions alone of community (Brint, 2001). Attempts made to find structural variables include: dense and demanding ties, social attachments and involvement in institutions, ritual occasions, small group size. Cultural variables include: perceptions of similarity with others, common beliefs in an idea system, moral order. Despite the issues he outlined, Brint (2001:8) attempted to offer a definition and outlined as the primary basis for relations in a community: affect, loyalty, shared values or personal involvement in the lives of others.

6.2.4 Online communities and publics

In the online space, community has clearly become the main label used to describe this arena. Faraj et al., (2011) reviewed this literature and suggested a suitable definition of online communities as,

“open collectives of dispersed individuals with members who are not necessarily known or identifiable and who share common interests, and these communities attend to both their individual and their collective welfare” (Sproull and Arriaga 2007, cited in Faraj et al., 2011: 3)

This definition begins to loosen the overemphasis on social ties and refers to a *collective*. To a certain extent, the question of whether online collectives should be classified as communities has been addressed many times and simply dismissed as not being helpful. The traditional ideal of community is anyway little more than a “pastoralist myth” (Wellman & Gulia, 1999). However, the image of community as being a core-periphery structure (Borgatti & Everett, 2000) is very present in the sociological and organisational theory literatures. In many conceptions, community is about a stable unit, one where social ties are well formed and identity is shared.

Within recent history of online research, the concept of community has been central, even if rhetorical as some have suggested the distinction is not really that useful (Preece, 2000). However, the logic of sharedness and core-periphery is especially evident in studies on innovation (Franke & Shah, 2003; Dahlander & Frederikson, 2011). It has recently been suggested that online communities need to be considered as more dynamic collectivities of activity, constantly in flux (Faraj et al., 2011). This thesis is an effort to consider how community and the alternative organising principle of publics may be represented in an online forum discussing

rumour about new products. Work on brand community (Muniz et al., 2006) develops rumour as a function of community, with a “shared consciousness of kind.” Shibutani (1966) discusses rumours as being debated and deliberated in a public. He explicitly distinguished the public from a community, referring to the latter as occupying a territorial unit. Both brand community and recent work in Economic Sociology on the construction of markets as fields have accepted communities as core-periphery structures. These entail a dense core of dedicated members, who form strong ties and maintain them through regular interaction. The brand community core is characterised as “retelling the brand narrative” (Muniz et al., 2006). Brand communities have been discussed as potentially being a source of innovation to firms (Fuller et al., 2008).

Rumours and speculation are deliberated in issue publics. The concept of a ‘public’ goes back to Tarde (1901) and Dewey (1927), though it has often overlapped with the ‘public sphere’ (Habermas, 1992). The critical distinction is that publics are temporary social collectivities that form around issues. This is in contrast to the public sphere, which is seen to be more stable. Habermas (1992) developed his public sphere around the model of romantic French Salons, who get together and rationally deliberate. The concept of a public here is a form of collective behaviour that is enabled by mass communication. People, likely not otherwise socially connected, gather in a discursive setting and interact around the issue. Although the *publics* concept has been present in the fields of communication studies and public relations, it has only recently been revived in cultural sociology by White (1995) and Mische (2009) and has not yet been discussed within organisation theory. I use the

publics concept to help illustrate a grouping of social interaction that occurs around new product speculation by a collectivity outside the individual firm.

What publics and community share is the concept of a platform. Information flow occurs in a public because of the platform of mass media. Flows in a community can be considered as occurring through social ties. Both support platforms, but the mechanisms are different. No known studies have considered the temporal and discursive structures of publics. Many studies have analytically considered community through a core-periphery structure (Borgatti & Everett, 1999).

The structure of community is important as a frame for a field, market and industry.

DiMaggio & Powell (1983: 148) laid out a field as

“organisations that, in the aggregate, constitute a recognised area of institutional life: key suppliers, resource and product consumers, regulatory agencies, and other organisations, unions, or boards of directors, that produce similar services or products.”

Institutional theory has maintained this strong interest in the community as a network of social or organisational actors. Fligstein (1996) and White (1981) maintained a strong focus on a market as mainly consisting of a network of producers. Population ecology studied networks of competing firms (Hannan & Freeman, 1977; Laumann et al., 1978). Social relations come to constitute social structures, which fields certainly stabilise around.

Irrespective of the issues with the definition of community, the public was always seen as being less cohesive and structured. In Table 6.1, I present a comparison that delineates some of the major differences between the two concepts. Much work in the online community domain has focused on the core-periphery structure (Jeppesen

& Lakhani, 2010; Dahlander & Frederiksen, 2011). Publics may have a structure, but it will be far less static and ephemeral around an event.

	<i>Community</i>	<i>Public</i>
<i>Bonding mechanisms</i>	Consciousness-of-kind; shared sense of identity;	Centred around a problematic situation
Composition	Once established, fairly steady structure	Varying stability through phases (latent, aware, active)
Boundaries	Set by core membership;	Set by event; limits of effective communication
Form of organisation	Can be organised, membership may not be fixed, but will have a stable core; inhabits particular territorial units	Not an organised group; does not have a fixed and easily identifiable membership
Structure	Core-periphery	Activists; Spectators
Bonds	Strong-ties in the core, Weak ties in periphery	Mostly weak ties

Table 6.1: Conceptually separating Community & Public

6.2.5 Method: Noun phrases in concept network analysis

Next, I discuss my methodological contributions. I contribute through developing a computational approach to text analysis by projecting concept networks. This method accounts for multiword noun phrases, has a fuzzy-string algorithm and uses modularity as a network measure. With an unsupervised machine learning linguistic algorithms to extract sentences and then noun phrases from them. I project a concept network through the co-occurrences of noun phrases in sentences. This method can be used to study meaning structures (Mohr, 1998). It was used in this dissertation to show a nomenclature stabilising structurally. It also speaks to the centrality of labels as cognitive referents and informs work on emergence of meaning structures

6.2.6 Method: Considering rumour and speculation as data

This dissertation also contributes to a pragmatist perspective of rumour in management and organisational theory. Rumour and speculation are clearly not to be afforded the same social status as facts, but they help to serve a role in social constructions in nascent markets. Rumour research has identified the differences in studying the process in a laboratory against a more naturalistic setting (Kapferer, 1990; Rosnow, 1991). This body of work on rumour may help to create a methodology to better inform firm interaction with external constituencies. Because shared schemas are socially determined (Dougherty, 1990; Rindova & Petkova, 2007), it seems troublesome to engage a minority of users in a directed laboratory as

individual prophets to speak on behalf of a larger cohort of consumers in a nascent market. An online community could very well be characterised as a social collectivity of interested and informed social actors who engage in public deliberation. They may not have the loyalties of a brand community, but can still end up generating discourses. This may well serve as conceptualising nascent product ontologies.

The stigma of rumours in popular discourse has led to a conceptualisation that they are always false and destructive. It is not surprising to see this given the context of societal efforts that stress scientific, objective and rational discourse. Intermediaries and experts play a major role in generating and certifying truth claims (Fine & DiFonzo, 2011). The benefits to orienting social relations around rational discourse are abundantly clear. However, as Shibutani (1966) suggests, there are pragmatic concerns that prevent people in context from only trading in verified objective propositions. This is especially evident in with a firm, where managers on a daily basis are pressed for time to make decisions amidst uncertainty and ambiguity. This echoes Mintzberg's (1989) contention that managers need to be aware of 'soft' unstructured information,

"Managers seem to cherish 'soft' information, especially gossip, hearsay, and speculation. Why? The reason is its timeliness; today's gossip may be tomorrow's fact. The manager who is not accessible for the telephone call informing him that the biggest customer was seen golfing with his main competitor may read about a dramatic drop in sales in the next quarterly report. But then it's too late." (Mintzberg, 1989: 13)

Arenas such as product publics can indeed be trading on potentially true rumours. Due to uncertainty in industry and innovation, the veracity of the information is not necessarily known. This class of information is a duality, and may even become a self-fulfilling truth (Merton, 1968). In high-technology markets, characterised by

heterogeneity of standards (Garud & Kumaraswamy, 1993; Sanchez & Mahoney, 1996) and competence-destroying discontinuities (Tushman & Anderson, 1986), uncertainty and timeliness matter. The interesting question for innovation studies is whether product rumours can, through their influence on the innovation process, become the basis for cognitive knowledge structures as expectations around a market. More important, they may be developing as propositions that converge into product ontologies.

6.2.7 Contributions to innovation studies

The core argument in this perspective is that innovation rumours develop through discourse and thus into the social shaping of innovations. Seen as *cheap talk*, this form of discourse is part of a sensemaking process where rumours filtered through a network of social relations (Fine & DiFonzo, 2011). However, it is not always clear how well formed social relations are. Particularly in a mass media such as on the Internet, a public may more appropriately capture this collective activity. This concept potentially captures the dynamic of early interaction that occurs between social groups attempting to attribute meanings to technical artefacts (Pinch & Bijker, 1997). While the interpretive flexibility is still high, there is considerable uncertainty over the definition of the innovation and so suggest that rumours play a role in the jostling for a dominant position. Although innovation rumours will have a low degree of formalisation, as they take on different constructions according to the conditions of ambiguous situation, they may be subject to immense deliberation.

People would prefer to have their news from legitimate sources (Shibutani, 1966). However, when there is no news to be found, rumours fill the gap and become a substitute for it. In the interaction between relevant social groups, we must recall this, as the contestation is not always itself formalised. Users, designers, and analysts seem to look within to derive expectations for the eventual dominant interpretation. Indeed their formation as social groups may be tied to reflexive consideration of innovation narratives. Product ontologies may begin to develop as the conditions of the situation make way for the emergence of rumours. As rumours can be *interpretive improvisations* (de Fleur, 1962), fresh perspectives may even result. When the firm signals direction through innovation disclosure events, this serves as a legitimising force that places pressure on the ambiguous situation, thus likely forcing rumours to intensify in the product rumour public. Finally, when the innovation disclosure event is staged, it will serve as a force to terminate the rumours.

Studying speculation and rumour like this also speaks to an epistemological shift in our consideration of information flow and structure. In the early days of rumour (Allport & Postman, 1947), rumour was studied by social psychologists who used the frame of a network to study information transmission. Similar to the *broken telephone game*, this considered information flowing along channels, much like an electrical signal traveling along a wire. This was criticised by Shibutani (1966) as being too restrictive and divorced from naturalistic settings. Instead, he chose to consider the model of a collective transaction. That is, the ephemeral natural of information flow in naturalistic situations is not linear in its diffusion, and it may well evolve as it moves. Unfortunately for Shibutani, he was working in a time where technology did not exist to record such transactions, nor analyse them with large scale linguistics. His studies

were then limited to historical case studies using secondary data. Social media presents an opportunity for researchers because the interactions are already recorded and transcribed in text. This permits us now to investigate old theory with new scales of data and sophisticated methods (Watts, 2004).

Rumours represent discussion of future states that people care about. The content is situated along shared schemas that are held by the social context. With rumours of a product innovation, the novel concepts contained within will create an information black market, mobilise social groups, and potentially challenge existing schemas (Clark, 1985) both within the firm and the online *product public*.

6.3 Limitations

As an exploratory study, this dissertation has several limitations. First, the scope was around a single product ontology emerging from tech blog discourse. The results may be due to the particular features of the phenomenon that only apply to this single case. Generalising from the *tablet* case may be difficult. Particularly because each product ontology may emerge somewhat differently. The tablet was rumoured to emerge for many years before the market settled. This enabled a *product public* to develop, through time and space to deliberate ideas. There were rumours about the Apple's tablet back in 2006, and they continued well on until 2010.

Furthermore, clearly Apple is a company that attracts tremendous interest from the mainstream media and social media. *Product publics* will likely not emerge in every industry. Extending Shibutani's conditions of rumour construction, I outline that *product publics* will emerge in online forums during situations of ambiguity and uncertainty. This may cover blogs in nascent markets, but more mundane markets such as industrial chemicals will not garner much public interest.

Another limitation is that of the method. Computational text analysis is fundamentally based on the simple projection of overlapping noun phrases. This is a *simple parsing* technique that finds the smallest possible noun phrase sequence. The advantage of simple projection is a commensurability across large text corpora. However, the weakness is a lossy projection of meaning. However, every discourse analysis method depends on explicit utterances. This will ignore contextual cues of meaning that are not constituted in an overlapping structure. However, there is a chance this

will at least be tracked by the method. Caution is stressed with this method, as the researcher should wisely operate it in a semi-automated fashion.

6.4 Conclusions

Simacova & Neyland (2008) reported on an ethnographic study of a firm attempting to launch a radical new product through strategic communication to external constituencies. This involved generating and transmitting future facing narratives. Their efforts were hampered by the simple fact that the firm was not in a position of power to dictate the direction of the developing discourse. This highlights the interorganisational environment that most firms face.

As research frames are shifting to embrace this social view of product development, there is also a disruptive force within the firm with jurisdictions over information. Most firms are setup with some departments more outward-facing than others. A public relations manager is not an engineer who can design and build products. However, increasingly products are constellations of attributes designed by outside suppliers. This emphasises Wang's (2009) work on innovation concepts as traveling across organisational boundaries more-so than material objects do. It remains to be seen whether social media will be taken by firms as a tool to further open porous boundaries, or a high risk channel that undermines their efforts to control information. There are arguments for firms being more and less open, and an ideal outcome is not clear.

However, the innovation community view (Van de Ven, 1993) does speak to the embeddedness of firms in fields and communities. This dissertation builds a further understanding of how product ontologies are developed in a space that is open for those who look. This speaks to a “semi-public” dynamic of a platform, similar to the contexts of Silicon Valley (Saxenian, 1994; Brown & Duguid, 2000) and the industrial districts of Manchester (Marshall, 1890).

From a practitioner standpoint, fitting an awareness of the product ontology development into a firm’s product development cycle could well be helpful. This is not to say that with these computational tools one can precisely track at what stage of development the product ontology is. However, it can give some indication of the level by which the meaning structure has consolidated and be a “needs forecasting laboratory” (von Hippel, 1988).

This study highlights an important question regarding social media data. When used for strategy, it may serve a slightly different purpose than marketing and will lead to jurisdictional struggles. This is a new information technology in that previously marketing departments handled the constituency of the consumer base, now an engineer or CEO can check a blog and see a narrative or issue easily without the need for a legitimate voice or reputation.

References

- Abbott, A. D. 1988. **The system of professions**. University Of Chicago Press.
- Abrahamson, E., & Fombrun, C. J. 1994. Macrocultures: Determinants and Consequences. **Academy of Management Review**, 19(4): 728–755.
- Aldrich, H. E., & Fiol, C. M. 1994. Fools Rush in? The Institutional Context of Industry Creation. **Academy of Management Review**, 19(4): 645–670.
- Allport, G. W., & Postman, L. 1947. **The Psychology of Rumor**. Henry Holt.
- Anand, N., & Peterson, R. A. 2000. When Market Information Constitutes Fields: Sensemaking of Markets in the Commercial Music Industry. **Organization Science**, 11(3): 270–284.
- Anderson, P., & Tushman, M. L. 1990. Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change. **Administrative Science Quarterly**, 35(4): 604–633.
- Adner, R. 2012. **The Wide Lens**. Portfolio Penguin.
- Ansari, S., & Phillips, N. 2011. Text Me! New Consumer Practices and Change in Organisational Fields. **Organization Science**, 22(6): 1579–1599.
- Armstrong, A., & Hagel, J., III. 1996. The Real Value of ON-LINE Communities. **Harvard Business Review**, 74(3): 134–141.
- Arora, A., Fosfuri, A., & Gambardella, A. 2001. Markets for Technology and their Implications for Corporate Strategy. **Industrial and Corporate Change**, 10(2): 419–451.
- Axelrod, R. M. 1976. **Structure of Decision: The Cognitive Maps of Political Elites** (1st ed.). Princeton Univ Pr.

- Baldwin, C., Hienerth, C., & Hippel, von, E. 2006. How user innovations become commercial products: A theoretical investigation and case study. **Research Policy**, 35(9): 1291–1313.
- Barr, P. S., Stimpert, J. L., & Huff, A. S. 1992. Cognitive Change, Strategic Action, and Organisational Renewal. **Strategic Management Journal**, 13: 15–36.
- Barabási, A.-L., & Frangos, J. 2002. **Linked: The New Science of Networks** (1st ed.). Perseus Books Group.
- Barney, J. 1991. Firm Resources and Sustained Competitive Advantage. **Journal of Management**, 17(1): 99–120.
- Belz, A. 2002. Learning grammars for noun phrase extraction by partition search.
- Berger, P. L., & Luckmann, T. 1967. **The Social Construction of Reality: A Treatise in the Sociology of Knowledge**. Anchor.
- Benjamin, B. A., & Podolny, J. M. 1999. Status, Quality, and Social Order in the California Wine Industry. **Administrative Science Quarterly**, 44(3): 563–589.
- Bhattacharya, S., & Ritter, J. R. 1983. Innovation and Communication: Signalling with Partial Disclosure. **The Review of Economic Studies**, 50(2): 331–346.
- Bijker, W. E. 1997. **Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change (Inside Technology)**. The MIT Press.
- Bijker, W. E., Hughes, T. P., & Pinch, T. (Eds.). 1987. **The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology**. The MIT Press.
- Bilton, N. 2009. What Would a 10-Inch Apple “iSlate” Look Like? **New York Times**. <http://bits.blogs.nytimes.com/2009/12/28/what-would-a-10-inch-islate-look-like/>.

- Bingham, C., & Kahl, S. 2012. The Process Of Schema Emergence: Assimilation, Deconstruction, Unitization And The Plurality Of Analogies. **Academy of Management Journal**.
- Bird, S., Klein, E., & Loper, E. 2009. **Natural Language Processing with Python** (1st ed.). O'Reilly Media.
- Blei, D. 2012. Probabilistic topic models. **Communications of the ACM**, 55(4): 77.
- Blumer, H. 1948. Public Opinion and Public Opinion Polling. **American Sociological Review**, 13(5): 542–549.
- Blumer, H. 1966. The mass, the public, and public opinion. In B. Berelson & M. Janowitz (Eds.), **Reader in public opinion and communication** (2nd ed.): 43–50. New York: Free Press.
- Boje, D. 2001. **Narrative Methods for Organisational & Communication Research (SAGE series in Management Research)** (1st ed.). Sage Publications Ltd.
- Boje, D. M., Oswick, C., & Ford, J. D. 2004. Introduction to Special Topic Forum: Language and Organization: The Doing of Discourse. **The Academy of Management Review**, 29(4): 571–577.
- Bordia, P., Jones, E., Gallois, C., Callan, V., & DiFonzo, N. 2006. Management Are Aliens!: Rumors and Stress during Organizational Change. **Group & Organization Management**, 31(5): 601.
- Bordia, P., & Rosnow, R. L. 1998. Rumor Rest Stops on the Information Highway Transmission Patterns in a Computer-Mediated Rumor Chain. **Human Communication Research**, 25(2): 163–179.
- Bordia, P., DiFonzo, N., & Chang, A. 1999. Rumor as group problem solving: Development patterns in informal computer-mediated groups. **Small Group Research**, 30(1): 8.
- Bordia, P., & DiFonzo, N. 2002. When social psychology became less social: Prasad and the history of rumor research. **Asian Journal of Social Psychology**, 5(1): 49–61.

- Bordia, P., & DiFonzo, N. 2004. Problem Solving in Social Interactions on the Internet: Rumor As Social Cognition. **Social Psychology Quarterly**, 67(1): 33–49.
- Bordia, P., DiFonzo, N., & Chang, A. 1999. Rumor as group problem solving: Development patterns in informal computer-mediated groups. **Small Group Research**, 30(1): 8.
- Borgatti, S. P., & Everett, M. 2000. Models of core-periphery structures. **Social Networks**, 21(4): 375–395.
- Borgatti, S. P., & Foster, P. 2003. The network paradigm in organisational research: A review and typology. **Journal of Management**, 29(6): 991.
- Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. 2009. Network Analysis in the Social Sciences. **Science**, 323(5916): 892–895.
- Bourdieu, P. 1977. **Outline of a Theory of Practice**. Cambridge University Press.
- Boudreau, K., & Lakhani, K. R. 2009. How to Manage Outside Innovation. **MIT Sloan Management Review**.
- Brandenburger, A. M., & Nalebuff, B. J. 1995. The right game: use game theory to shape strategy. **Harvard Business Review**.
- Breiger, R. L. 2000. A tool kit for practice theory. **Poetics**, 27(2–3): 91–115.
- Brint, S. 2001. Gemeinschaft Revisited: A Critique and Reconstruction of the Community Concept. **Sociological Theory**, 19(1): 1–23.
- Brown, J. S., & Duguid, P. 2000. Mysteries of the region: knowledge dynamics in Silicon Valley. In C. M. Lee (Ed.), **The Silicon Valley Edge: A Habitat for Innovation and Entrepreneurship**: 16–45. Stanford University Press.
- Brown, J. S., & Duguid, P. 2002. **Social Life Of Information**. Harvard Business School Press.

- Brown, S. L., & Eisenhardt, K. 1995. Product Development: Past Research, Present Findings, and Future Directions. **The Academy of Management Review**, 20(2): 343–378.
- Burt, R. S. 2004. Structural Holes and Good Ideas. **American Journal of Sociology**, 110(2): 349–399.
- Burt, R. S. 2008. Gossip and Reputation. In P. Lièvre & M. Lecoutre (Eds.), **Management Et Réseaux Sociaux : Ressource Pour L'action Ou Outil de Gestion?**: 27–42. Paris: Hermès science publications.
- Bussmann, H. 1996. **Routledge Dictionary of Language and Linguistics**.
- Carley, K. M. 1986. An approach for relating social structure to cognitive structure. **The Journal of Mathematical Sociology**, 12(2): 137–189.
- Carley, K. M. 1993. Coding choices for textual analysis: A comparison of content analysis and map analysis. **Sociological methodology**, 23: 75–126.
- Carley, K. M. 1994. Extracting culture through textual analysis. **Poetics**, 22(4): 291–312.
- Carley, K. M. 1997a. Extracting team mental models through textual analysis. **Journal of Organisational Behavior**, 18(S1): 533–558.
- Carley, K. M. 1997b. Network text analysis: The network position of concepts. In C. W. Roberts (Ed.), **Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts**. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Carley, K. M., & Kaufer, D. S. 1993. Semantic Connectivity: An Approach for Analysing Symbols in Semantic Networks. **Communication Theory**, 3(3): 183–213.
- Carley, K. M., & Palmquist, M. 1992. Extracting, Representing, and Analysing Mental Models. **Social Forces**, 70(3): 601–636.

- Carley, K. M. 2002. Computational organization science: A new frontier. **Proceedings of the National Academy of Sciences of the United States of America**, 99(Suppl 3): 7257–7262.
- Carlile, P. R. 2002. A pragmatic view of knowledge and boundaries: boundary objects in new product development. **Organization Science**, 13(4): 442–455.
- Chang, J., Boyd-Graber, J., Wang, C., Gerrish, S., & Blei, D. 2009. Reading tea leaves: How humans interpret topic models. **Presented at the Proceedings of the Neural Information Processing Systems**.
- Cheng, A.-S., Fleischmann, K., Wang, P., & Oard, D. W. 2008. Advancing social science research by applying computational linguistics. **Proceedings of the American Society for Information Science and Technology**, 45(1): 1–12.
- Chesbrough, H., Vanhaverbeke, W., & West, J. 2006. **Open innovation: Researching a new paradigm**. Oxford University Press, USA.
- Chomsky, N. 1956. Three models for the description of language. **IEEE Transactions on Information Theory**, 2(3): 113–124.
- Christensen, C. 1997. **The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail** (1st ed.). Harvard Business Review Press.
- Clark, K. B. 1985. The interaction of design hierarchies and market concepts in technological evolution. **Research Policy**, 14(5): 235–251.
- Clark, D. 2011. Samsung Cites Sci-Fi Classic in Attacking Apple Patent. **Wall Street Journal**. <http://blogs.wsj.com/digits/2011/08/23/samsung-cites-sci-fi-classic-in-attacking-apple-patent/>.
- Clark, K. B., & Fujimoto, T. 1991. **Product Development Performance: Strategy, Organization, and Management in the World Auto Industry**. Harvard Business Press.
- Cohen, A. P. 1985. **Symbolic Construction of Community (Key Ideas)** (New Ed.). Routledge.

- Collins, H. 2004. **Gravity's Shadow: The Search for Gravitational Waves** (1st ed.). University Of Chicago Press.
- Contractor, N., Wasserman, S., & Faust, K. 2006. Testing Multitheoretical, Multilevel Hypotheses About Organisational Networks: An Analytic Framework And Empirical Example. **Academy of Management Review**, 31(3): 681–703.
- Cooper, R. G. 2001. **Winning at New Products: Accelerating the Process from Idea to Launch**, Third Edition (3rd ed.). Basic Books.
- Corman, S. R., Kuhn, T., Mcphee, R. D., & Dooley, K. J. 2002. Studying Complex Discursive Systems. **Human Communication Research**, 28(2): 157–206.
- Csardi, G., & Nepusz, T. 2006. The igraph software package for complex network research. **InterJournal**.
- d'Aspremont, C., Bhattacharya, S., & Gérard-Varet, L.-A. 2000. Bargaining and Sharing Innovative Knowledge. **The Review of Economic Studies**, 67(2): 255–271.
- Dacin, M. T., Ventresca, M., & Beal, B. D. 1999. The Embeddedness of Organizations: Dialogue & Directions. **Journal of Management**, 25(3): 317–356.
- Daelemans, W., Buchholz, S., & Veenstra, J. 1999. Memory-based shallow parsing. **Presented at the Proceedings of CoNLL**.
- Dahlander, L., & Frederiksen, L. 2011. The Core and Cosmopolitans: A Relational View of Innovation in User Communities. **Organization Science**, 23(4): 988–1007.
- Dahlander, L., Frederiksen, L., & Rullani, F. 2008. Online Communities and Open Innovation. **Industry & Innovation**, 15(2): 115–123.
- Dasgupta, P., & Stiglitz, J. 1980. Uncertainty, Industrial Structure, and the Speed of R&D. **The Bell Journal of Economics**, 11(1): 1–28.

- David, P. A. 1992. Heroes, Herds and Hysteresis in Technological History: Thomas Edison and “The Battle of the Systems” Reconsidered. **Industrial and Corporate Change**, 1(1): 129–180.
- David, P. A., & Shapiro, J. 2008. Community-based production of open-source software- What do we know about the developers who participate? **Information Economics and Policy**, 20: 1–35.
- Davidson, E., & Vaast, E. 2009. Tech talk: An investigation of blogging in technology innovation discourse. **Professional Communication, IEEE Transactions on**, 52(1): 40–60.
- de Fleur, M. 1962. Mass Communication and The Study of Rumor*. **Sociological Inquiry**, 32(1): 51–70.
- De Smedt, T., & Daelemans, W. 2012. Pattern for Python. **The Journal of Machine Learning Research**, 98888: 2063–2067.
- Deephouse, D. L., & Heugens, P. P. 2009. **Journal of business ethics**. 86(4)
- Dewey, J. 1927. **The public and its problems**. Swallow Press.
- DiFonzo, N., & Bordia, P. 1998. A tale of two corporations: Managing uncertainty during organisational change. **Human Resource Management**, 37.
- DiFonzo, N., & Bordia, P. 2000. How top PR professionals handle hearsay: corporate rumors, their effects, and strategies to manage them. **Public Relations Review**, 26(2): 173–190.
- DiMaggio, P. 1991. **Constructing and Organisational Field as a Professional Project: U.S. Art Museums, 1920–1940. The New Institutionalism and Organisational Analysis**: 267–292. University of Chicago Press.
- DiMaggio, P. 1997. Culture and Cognition. **Annual Reviews in Sociology**, 23: 263–287.

- DiMaggio, P., & Louch, H. 1998. Socially Embedded Consumer Transactions: For What Kinds of Purchases Do People Most Often use Networks? **American Sociological Review**, 63(5): 619–637.
- DiMaggio, P., & Powell, W. W. 1983. The iron cage revisited: Institutional isomorphism and collective rationality in organisational fields. **American Sociological Review**, 147–160.
- Doerfel, M. L., & Connaughton, S. L. 2009. Semantic networks and competition: Election year winners and losers in U.S. televised presidential debates, 1960–2004. **Journal of the American Society for Information Science and Technology**, 60(1): 201–218.
- Dougherty, D. 1990. Understanding New Markets for New Products. **Strategic Management Journal**, 11: 59–78.
- Dougherty, D. 1992. Interpretive Barriers to Successful Product Innovation in Large Firms. **Organization Science**, 3(2): 179–202.
- Douglas, M. 1987. **How Institutions Think**. London: Routledge.
- Droge, C., Stanko, M. A., & Pollitte, W. A. 2010. Lead Users and Early Adopters on the Web: The Role of New Technology Product Blogs*. **Journal Of Product Innovation Management**, 27(1): 66–82.
- Durkheim, E. 1915. **The elementary forms of religious life**. Free Press.
- Durkheim, E. 1997. **Suicide**. (G. Simpson, Ed. J. A. Spaulding, Tran.). Free Press.
- Dwyer, P. 2012. An approach to measuring influence and cognitive similarity in computer-mediated communication. **Computers in Human Behavior**, 28(2): 540–551.
- Dybwad, B. 2010. 9 Upcoming Tablet Alternatives to the Apple iPad. **Mashable**.<http://mashable.com/2010/01/27/9-upcoming-tablet-alternatives-to-the-apple-ipad/>, January 2013.
- Dym, C. L., & Little, P. 2000. **Engineering design: A project-based approach**. John Wiley & Sons, Inc.

- Eisenhardt, K. 1989. Making Fast Strategic Decisions in High-Velocity Environments. **Academy of Management Journal**, 32(3): 543–576.
- Ernst, H. 2002. Success Factors of New Product Development: A Review of the Empirical Literature. **International Journal of Management Reviews**, 4(1): 1–40.
- Faraj, S., & Johnson, S. L. 2010. Network Exchange Patterns in Online Communities. **Organization Science**, 22(6): 1464–1480.
- Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. 2011. Knowledge collaboration in online communities. **Organization Science**, 22(5): 1224–1239.
- Fauchart, E., & Hippel, von, E. 2008. Norms-Based Intellectual Property Systems: The Case of French Chefs. **Organization Science**, 19(2): 187–201.
- Faulkner, P., & Runde, J. 2009. On the Identity of Technological Objects and User Innovations in Function. **Academy of Management Review**, 34(3): 442–462.
- Fine, G. A. 1995. **A Second Chicago School?: The Development of a Postwar American Sociology**. (G. A. Fine, Ed.). University Of Chicago Press.
- Fine, G. A. 2010. The Sociology of the Local: Action and its Publics*. **Sociological Theory**, 28(4): 355–376.
- Fine, G. A., & DiFonzo, N. 2011. Uncertain Knowledge. **Contexts**, 10(3): 16–21.
- Fine, G. A., & Ellis, B. 2010. **The Global Grapevine: Why Rumors of Terrorism, Immigration, and Trade Matter**. Oxford University Press, USA.
- Fine, G. A., & Harrington, B. 2004. Tiny Publics: Small Groups and Civil Society. **Sociological Theory**, 22(3): 341–356.
- Fisher, D., Smith, M., & Welser, H. T. 2006. You Are Who You Talk To: Detecting Roles in Usenet Newsgroups: 59b–59b. Presented at the **System Sciences, 2006. HICSS '06. Proceedings of the 39th Annual Hawaii International Conference on, IEEE**.

- Fletcher. 2010. Best blogs of 2010. **TIME Magazine**. http://www.time.com/time/specials/packages/article/0,28804,1999770_1999761_1999863,00.html, March 2011.
- Fligstein, N. 1996. Markets as Politics: A Political-Cultural Approach to Market Institutions. **American Sociological Review**, 61(4): 656–673.
- Fligstein, N. 2002. **The Architecture of Markets: An Economic Sociology of Twenty-First-Century Capitalist Societies**. Princeton University Press.
- Fligstein, N., & Dauter, L. 2007. The Sociology of Markets. **Annual Reviews in Sociology**, 33(1): 105–128.
- Forbes, D. P., & Kirsch, D. A. 2011. The study of emerging industries: Recognizing and responding to some central problems. **Journal of Business Venturing**, 26(5): 589–602.
- Foster, R. N. 1985. Timing technological transitions. **Technology in society**, 7(2–3): 127–141.
- Franke, N., & Hippel, E. V. 2003. Satisfying heterogeneous user needs via innovation toolkits: the case of Apache security software. **Research Policy**, 32(7): 1199–1215.
- Franke, N., & Shah, S. 2003. How communities support innovative activities: an exploration of assistance and sharing among end-users. **Research Policy**, 32(1): 157–178.
- Franzosi, R. 1994. From words to numbers: A set theory framework for the collection, organization and analysis of narrative data. **Sociological methodology**, 24: 105–136.
- Freeman, L. C. 1977. A Set of Measures of Centrality Based on Betweenness. **Sociometry**, 40(1): 35–41.
- Frickel, S., & Moore, K. 2006. **The New Political Sociology of Science**. Univ of Wisconsin Press.
- Friedel, R. 1994. The History of the Zipper. **American Heritage of Invention & Technology**, 10(1): 8-16

- Fujimura, J. 1997. **Crafting Science: A Sociohistory of the Quest for the Genetics of Cancer**. Harvard University Press.
- Fuller, J., Bartl, M., Ernst, H., & Mühlbacher, H. 2006. Community based innovation: How to integrate members of virtual communities into new product development. **Electronic Commerce Research**, 6(1): 57–73.
- Fuller, J., Matzler, K., & Hoppe, M. 2008. Brand community members as a source of innovation: 608–619. **Presented at the Journal of Product Innovation Management**.
- Gans, J., & Stern, S. 2003. The product market and the market for “ideas”: commercialization strategies for technology entrepreneurs. **Research Policy**, 32(2): 333–350.
- Garud, R., & Kumaraswamy, A. 1993. Changing Competitive Dynamics in Network Industries: An Exploration of Sun Microsystems' Open Systems Strategy. **Strategic Management Journal**, 14(5): 351–369.
- Garud, R., & Rappa, M. A. 1994. A Socio-Cognitive Model of Technology Evolution: The Case of Cochlear Implants. **Organization Science**, 5(3): 344–362.
- Garud, R., Lant, T., & Schildt, H. A. 2008. A discursive approach to entrepreneurship in the emergence of organisational fields. **Working paper**
- Geertz, C. 1973. **The Interpretation Of Cultures (Basic Books Classics)**. Basic Books.
- Geroski, P. A. 2003. **The Evolution of New Markets**. Oxford University Press, USA.
- Giddens, A. 1984. **The Constitution of Society**. Los Angeles: University of California Press.
- Gilson, R. J. 1996. The Fading Boundaries of the Firm: The Role of Information and Communication Technology: Comment. **Journal of Institutional and Theoretical Economics**, 152(1): 287–294.

- Gieryn, T. 1996. Policing STS: A Boundary-Work Souvenir from the Smithsonian Exhibition on "Science in American Life." **Science, Technology & Human Values**, 21(1): 100–115.
- Goffman, E. 1959. **The Presentation of Self in Everyday Life** (1st ed.). Anchor.
- Goffman, E. 1974. **Frame analysis: An essay on the organization of experience**. Harvard University Press.
- Gonzalez-Bailon, S., Kaltenbrunner, A., & Banchs, R. E. 2010. The structure of political discussion networks: a model for the analysis of online deliberation. **Journal of Information Technology**, 25(2): 230–243.
- Goode, E. 1992. **Collective behavior**. Wadsworth Publishing.
- Granovetter, M. S. 1973. The Strength of Weak Ties. **American Journal of Sociology**, 78(6): 1360–1380.
- Granovetter, M. 1985. Economic action and social structure: the problem of embeddedness. **American Journal of Sociology**, 91(3): 481.
- Grodal, S. 2007. **The emergence of a new organisational field—Labels, meaning and emotions in nanotechnology**. Stanford University.
- Grosz, B. J., Weinstein, S., & Joshi, A. K. 1995. Centering: a framework for modeling the local coherence of discourse. **Computational linguistics**, 21(2): 203–225.
- Guiltinan, J. P. 1999. Launch Strategy, Launch Tactics, and Demand Outcomes. **Journal Of Product Innovation Management**, 16(6): 509–529.
- Habermas, J. 1992. **The structural transformation of the public sphere: An inquiry into a category of bourgeois society**. Polity Press.
- Hales, C., & Gooch, S. 2004. **Managing Engineering Design** (2nd ed.). Springer.
- Hannan, M. T., & Carroll, G. R. 1995. Theory Building and Cheap Talk About Legitimation: Reply to Baum and Powell. **American Sociological Review**, 60(4): 539–544.

- Hannan, M. T., & Freeman, J. 1977. The population ecology of organizations. **American Journal of Sociology**, 82(5): 929–964.
- Hargadon, A. 2006. Bridging old worlds and building new ones: towards a microsociology of creativity. **Creativity and innovation in organisational teams (Lea's Organization and Management)**, 199–216.
- Hargadon, A. B., & Douglas, Y. 2001. When Innovations Meet Institutions: Edison and the Design of the Electric Light. **Administrative Science Quarterly**, 46(3): 476.
- Hill, V., & Carley, K. M. 1999. An approach to identifying consensus in a subfield: The case of organisational culture. **Poetics**, 27(1): 1–30.
- Hippel, von, E. 1986. Lead Users: A Source of Novel Product Concepts. **Management Science**, 32(7): 791–805.
- Hippel, von, E. 1988. **The sources of innovation**. Oxford: Oxford University Press
- Hippel, von, E. 2001. Learning from open-source software. **MIT Sloan Management Review**, 82–86.
- Hippel, von, E. 2005. **Democratizing Innovation**. The MIT Press.
- Hippel, von, E., & Katz, R. 2002. Shifting Innovation to Users via Toolkits. **Management Science**, 48(7): 821–833.
- Hirsch, P. M. 1997. Sociology without Social Structure: Neoinstitutional Theory Meets Brave New World. **American Journal of Sociology**, 102(6): 1702–1723.
- Hodge, R., & Kress, G. 1988. **Social Semiotics**. Cornell University Press.
- Hoffman, A. J. 1999. Institutional Evolution and Change: Environmentalism and the U.S. Chemical Industry. **Academy of Management Journal**, 42(4): 351–371.
- Jacobs, J. 1992. **The Death and Life of Great American Cities**. Vintage.

- Jeppesen, L. B. 2005. User Toolkits for Innovation: Consumers Support Each Other. **Journal Of Product Innovation Management**, 22(4): 347–362.
- Jeppesen, L. B., & Frederiksen, L. 2006. Why Do Users Contribute to Firm-Hosted User Communities? The Case of Computer-Controlled Music Instruments. **Organization Science**, 17(1): 45–63. <http://orgsci.journal.informs.org/content/17/1/45>.
- Jeppesen, L. B., & Lakhani, K. R. 2010. Marginality and Problem-Solving Effectiveness in Broadcast Search. **Organization Science**, 21(5): 1016–1033.
- Jeppesen, L. B., & Molin, M. J. 2003. Consumers as Co-developers: Learning and Innovation Outside the Firm. **Technology Analysis & Strategic Management**, 15(3): 363–383.
- Johnson, S. 2011. **Where Good Ideas Come From: The Natural History of Innovation** (Reprint.). Riverhead Trade.
- Kahney, L. 2006. **The Cult of Mac (Paperback edition)**. No Starch Press.
- Kaltenbrunner, A., & Gonzalez-Bailon, S. 2009. Communities on the web: Mechanisms underlying the emergence of online discussion networks: **Proceedings of the WebSci'09: Society On-Line**, Information Science Publishing.
- Kapferer, J.N. 1990. **Rumors: uses, interpretations, and images**. Transaction Pub.
- Kaplan, S. 2010. Framing Contests: Strategy Making Under Uncertainty. **Organization Science**, 19(5): 729–752.
- Kaplan, S. 2011. Research in Cognition and Strategy: Reflections on Two Decades of Progress and a Look to the Future. **Journal of Management Studies**, 48(3): 665–695.
- Kaplan, S., & Tripsas, M. 2008. Thinking about technology: Applying a cognitive lens to technical change. **Research Policy**, 37(5): 790–805.

- Kaplan, S., & Vakili, K. 2012. Identifying Breakthroughs: Cognitive vs. Economic. **Working paper**.
- Katz, R. 1982. The effects of group longevity on project communication and performance. **Administrative Science Quarterly**, 27(1): 81–104.
- Katz, R., & Tushman, M. L. 1981. An investigation into the managerial roles and career paths of gatekeepers and project supervisors in a major R & D facility. **R&D Management**, 11(3): 103–110.
- Kennedy, M. T. 2005. Behind the one-way mirror: Refraction in the construction of product market categories. **Poetics**, 33(3-4): 201–226.
- Kennedy, M. T. 2008. Getting Counted: Markets, Media, and Reality. **American Sociological Review**, 73(2): 270–295.
- Kennedy, M. T., Lo, J. Y.-C., & Lounsbury, M. 2010. Category currency: The changing value of conformity as a function of ongoing meaning construction. **Research in the Sociology of Organizations**, 31: 369–397.
- Khurana, A., & Rosenthal, S. R. 1998. Towards Holistic “Front Ends” In New Product Development. **Journal Of Product Innovation Management**, 15(1): 57–74.
- Kilduff, M., Tsai, W., & Hanke, R. 2006. A Paradigm Too Far? A Dynamic Stability Reconsideration Of The Social Network Research Program. **Academy of Management Review**, 31(4): 1031–1048.
- Kilgarriff, A., & Grefenstette, G. 2003. Introduction to the Special Issue on the Web as Corpus. **Computational linguistics**, 29(3): 333–347.
- Kirchner, C., & Mohr, J. W. 2010. Meanings and relations: An introduction to the study of language, discourse and networks. **Poetics**, 38(6): 555–566.
- Klein, D., & Manning, C. D. 2003. Accurate unlexicalized parsing, **Proceedings of the 41st Annual Meeting on Association for Computational Linguistics** - 11: 423–430.

- Koenig, F. 1985. **Rumor in the marketplace**. Auburn House.
- Kozinets, R. V. 2002. The field behind the screen: using netnography for marketing research in online communities. **Journal of Marketing Research**, 39(1): 61–72. <http://www.jstor.org/stable/1558584>.
- Krackhardt, D. 1987. Cognitive social structures. **Social Networks**, 9(2): 109–134.
- Kreimer, A., Borenstein, E., Gophna, U., & Ruppín, E. 2008. The evolution of modularity in bacterial metabolic networks. **Proceedings of the National Academy of Sciences**, 105(19): 6976–6981.
- Krippendorff, K. H. 2003. **Content Analysis: An Introduction to Its Methodology** (2nd ed.). Sage Publications, Inc.
- Kuhn, T. 1970. **The Structure of Scientific Revolutions**. Chicago: University of Chicago Press.
- Kuhn, T. 2008. A Communicative Theory of the Firm: Developing an Alternative Perspective on Intra-organisational Power and Stakeholder Relationships. **Organization Studies**, 29(8-9): 1227–1254.
- Lakhani, K. R., & Hippel, von, E. 2003. How open source software works: “free” user-to-user assistance. **Research Policy**, 32(6): 923–943.
- Lakhani, K. R., & Wolf, R. 2003. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. **papers.ssrn.com**.
- Latour, B. 1987. **Science in action: how to follow scientists and engineers through society**. Harvard University Press.
- Lawrence, P. R., & Lorsch, J. W. 1967. **Organization and Environment: Managing Differentiation and Integration (Harvard Business School Classics)**. Harvard Business Press.

- Lawrence, T. B., & Phillips, N. 2004. From Moby Dick to Free Willy: Macro-Cultural Discourse and Institutional Entrepreneurship in Emerging Institutional Fields. **Organization**, 11(5): 689–711.
- Laumann, E. O., Galaskiewicz, J., & Marsden, P. V. 1978. Community structure as interorganizational linkages. **Annual Reviews in Sociology**, 4: 455–484.
- Leicht, E., & Newman, M. 2008. Community Structure in Directed Networks. **Physical Review Letters**, 100(11).
- Lerner, J., & Tirole, J. 2002. Some simple economics of open source. **Journal of Industrial Economics**, 197–234.
- Lilien, G., Morrison, P., Searls, K., Sonnack, M., & Hippel, von, E. 2002. Performance assessment of the lead user idea-generation process for new product development. **Management Science**, 1042–1059.
- Lippmann, W. 1925. **The phantom public: A sequel to “Public Opinion.”** Macmillan.
- Lohr, S. 2002. It’s a Tablet. It’s a Notebook. From Microsoft, a New Hybrid. **New York Times**. <http://www.nytimes.com/2002/06/25/technology/25SOFT.html>.
- Lounsbury, M., & Glynn, M. A. 2001. Cultural entrepreneurship: stories, legitimacy, and the acquisition of resources. **Strategic Management Journal**, 22(6-7): 545–564.
- Lounsbury, M., & Ventresca, M. 2003. The New Structuralism in Organisational Theory. **Organization**, 10(3): 457–480.
- Lounsbury, M., Ventresca, M., & Hirsch, P. M. 2003. Social movements, field frames and industry emergence: a cultural–political perspective on US recycling. **Socio-Economic Review**, 1(1): 71–104.
- Luthje, C., & Herstatt, C. 2004. The Lead User method: an outline of empirical findings and issues for future research. **R&D Management**, 34(5): 553–568.

- Luthje, C., Herstatt, C., & Hippel, von, E. 2005. User-innovators and “local” information: The case of mountain biking. **Research Policy**, 34(6): 951–965.
- Luthje, C., Herstatt, C., & Hippel, von, E. 2005. User-innovators and “local” information: The case of mountain biking. **Research Policy**, 34(6): 951–965.
- Manning, C. D., & Schuetze, H. 1999. **Foundations of Statistical Natural Language Processing** (1st ed.). The MIT Press.
- Manning, C. D., Raghavan, P., & Schütze, H. 2008. **Introduction to information retrieval** (1st ed.). Cambridge: University Press Cambridge.
- March, J. G. 1978. Bounded rationality, ambiguity, and the engineering of choice. **The Bell Journal of Economics**, 9(2): 587–608.
- March, J. G., & Simon, H. A. 1958. **Organizations**. Wiley.
- Marshall, A. 1890. **Principles of economics**. Macmillan and Company.
- McGahan, A. M., & Porter, M. E. 1997. How Much Does Industry Matter, Really? **Strategic Management Journal**, 18: 15–30.
- McPhail, C. 1991. **The Myth of the Madding Crowd (Social Institutions and Social Change)**. Aldine Transaction.
- Merton, R. K. 1968. **Social Theory and Social Structure** (1968th ed.). Free Press.
- Merton, R. K. 1979. **The sociology of science**. University Of Chicago Press.
- Meyer, A. D., Gaba, V., & Colwell, K. A. 2005. Organizing Far from Equilibrium: Nonlinear Change in Organizational Fields. **Organization Science**, 16(5): 456–473.
- Microsoft. 2001. Gates Showcases Tablet PC, Xbox at COMDEX; Says New “Digital Decade” Technologies Will Transform How We Live. <http://www.microsoft.com/en-us/news/press/2001/Nov01/11-11Comdex2001KeynotePR.aspx>, Microsoft.

- Mintzberg, H. 1989. **Mintzberg on Management**. Free Press.
- Mische, A. 2009. **Partisan Publics: Communication and Contention across Brazilian Youth Activist Networks (Princeton Studies in Cultural Sociology)**. Princeton University Press.
- Mische, A., & Pattison, P. 2000. Composing a civic arena: Publics, projects, and social settings. **Poetics**, 27(2-3): 163–194.
- Moenaert, R., Caeldries, F., Lievens, A., & Wauters, E. 2000. Communication flows in international product innovation teams. **Journal Of Product Innovation Management**, 17(5): 360–377.
- Mohr, J. W. 1998. Measuring Meaning Structures. **Annual Reviews in Sociology**, 24 IS -: 345–370.
- Monge, P. R., & Contractor, N. 2003. **Theories of Communication Networks**. Oxford University Press, USA.
- Monge, P. R., Cozzens, M. D., & Contractor, N. 1992. Communication and Motivational Predictors of the Dynamics of Organisational Innovation. **Organization Science**, 3(2): 250–274.
- Morrison, P., Roberts, J., & Midgley, D. 2004. The nature of lead users and measurement of leading edge status. **Research Policy**, 33(2): 351–362.
- Mowery, D., & Rosenberg, N. 1979. The influence of market demand upon innovation: a critical review of some recent empirical studies. **Research Policy**, 8(2): 102–153.
- Muller, P., & Pénin, J. 2005. Why do firms disclose knowledge and how does it matter? **Journal of evolutionary economics**, 16(1-2): 85–108.
- Munir, K. A., & Phillips, N. 2002. The concept of industry and the case of radical technological change. **The Journal of High Technology Management Research**, 13(2): 279–297.

- Munir, K. A., & Phillips, N. 2005. The Birth of the “Kodak Moment”: Institutional Entrepreneurship and the Adoption of New Technologies. **Organization Studies**, 26(11): 1665–1687.
- Muniz, A., Jr, & O'Guinn, T. 1995. Brand Community and the Sociology of Brands. **Advances in Consumer Research**, 23.
- Muniz, A., Jr, & O'Guinn, T. 2001. Brand Community. **Journal of Consumer Research**, 27(4): 412–432.
- Muniz, A., Jr, & Schau, H. 2005. Religiosity in the abandoned Apple Newton brand community. **Journal of Consumer Research**, 31(4): 737–747.
- Muniz, A., Jr, O'Guinn, T., & Fine, G. A. 2006. Rumor in brand community. In D. A. Hantula (Ed.), **Advances in social & organisational psychology**: 227–247. Erlbaum.
- Murray, F., & O'Mahony, S. 2007. Exploring the Foundations of Cumulative Innovation: Implications for Organization Science. **Organization Science**, 18(6): 1006–1021.
- Nambisan, S., & Baron, R. A. 2010. Different Roles, Different Strokes: Organizing Virtual Customer Environments to Promote Two Types of Customer Contributions. **Organization Science**, 21(2): 554–572.
- NPD DisplaySearch. 2013. Tablet PC Market Forecast to Surpass Notebooks in 2013, NPD DisplaySearch Reports. http://www.displaysearch.com/cps/rde/xchg/displaysearch/hs.xsl/130107_tablet_pc_market_forecast_to_surpass_notebooks_in_2013.asp.
- Oard, D. W. 2009. A Whirlwind Tour of Automated Language Processing for the Humanities and Social Sciences. **Working Together or Apart: Promoting the Next Generation of Digital Scholarship, Council on Library and Information Resources**, 34–42.
- O'Mahony, S. 2003. Guarding the commons: how community managed software projects protect their work. **Research Policy**, 32(7): 1179–1198.
- O'Mahony, S., & Bechky, B. A. 2008. Boundary Organizations: Enabling Collaboration among Unexpected Allies. **Administrative Science Quarterly**, 53(3): 422–459.

- Ocasio, W. 1997. Towards an attention-based view of the firm. **Strategic Management Journal**, 18: 187–206.
- Olson, D. 2008. Ethical aspects of web log data mining. **International Journal of Information Technology and Management**, 7(2): 190–200.
- Orlikowski, W. J., & Gash, D. 1994. Technological frames: making sense of information technology in organizations. **ACM Transactions on Information Systems (TOIS)**, 12(2): 174–207.
- Pachucki, M. A., & Breiger, R. L. 2010. Cultural Holes: Beyond Relationality in Social Networks and Culture. **Annual Reviews in Sociology**, 36(1): 205–224.
- Petroski, H. 1992. **The evolution of useful things**. Vintage.
- Pfeffer, J., & Salancik, G. R. 2003. **The External Control of Organizations**. Stanford University Press.
- Phillips, N., Lawrence, T. B., & Hardy, C. 2004. Discourse and Institutions. **The Academy of Management Review**, 29(4): 635–652.
- Phillips, N. & Hardy, C. 2002. **Discourse Analysis: Investigating Processes of Social Construction**. (N. Phillips, C. Hardy, Eds.) (1st ed.). Sage Publications, Inc.
- Pinch, T., & Bijker, W. E. 1987. The social construction of facts and artifacts. **The MIT Press, Cambridge, MA**, 17–50.
- Pinker, S. 2007. **The Language Instinct: How the Mind Creates Language (P.S.)** (third edition.). Harper Perennial Modern Classics.
- Podolny, J. M. 2001. Networks as the Pipes and Prisms of the Market. **American Journal of Sociology**, 107(1): 33–60.
- Polanyi, M. 1967. **The Tacit Dimension**. London: Routledge.

- Pollach, I. 2012. Taming Textual Data: The Contribution of Corpus Linguistics to Computer-Aided Text Analysis. **Organizational Research Methods**, 15(2): 263–287.
- Pollock, N., & Williams, R. 2011. Who decides the shape of product markets? The knowledge institutions that name and categorise new technologies. **Information and Organization**, 21(4): 194–217.
- Pollock, T. G., & Rindova, V. P. 2003. Media Legitimation Effects in the Market for Initial Public Offerings. **Academy of Management Journal**, 46(5): 631–642.
- Porac, J., & Thomas, H. 1990. Taxonomic Mental Models in Competitor Definition. **Academy of Management Review**, 15(2): 224–240.
- Porac, J., Thomas, H., Wilson, F., & Paton, D. 1995. Rivalry and the Industry Model of Scottish Knitwear Producers. **Administrative Science Quarterly**, 40(2): 203–227.
- Porac, J., & Rosa, J. A. 1996. Rivalry, industry models, and the cognitive embeddedness of the comparable firm. In J. C. Baum & J. E. Dutton (Eds) **Advances in strategic management**. 363–88 Greenwich, CT: JAI Press.
- Porac, J., Thomas, H., & Baden-Fuller, C. 1989. Competitive Groups As Cognitive Communities: The Case Of Scottish Knitwear Manufacturers*. **Journal of Management Studies**, 26(4): 397–416.
- Porac, J., Thomas, H., & Baden-Fuller, C. 2011. Competitive Groups as Cognitive Communities: The Case of Scottish Knitwear Manufacturers Revisited. **Journal of Management Studies**, 48(3): 646–664.
- Porac, J. F., Ventresca, M. J., & Mishina, Y. 2002. Interorganizational cognition and interpretation. In J. C. Baum (Ed.), **Companion to organizations**: 579–598: London: Blackwell.
- Porter, M. E. 1980. **Competitive strategy**. New York: Free Press.

- Pound, J., & Zeckhauser, R. 1990. Clearly heard on the street: The effect of takeover rumors on stock prices. **Journal of Business**, 63(3): 291–308. <http://www.jstor.org/stable/2353151>.
- Powell, W. W. 1990. Neither market nor hierarchy: Network forms of Organization. **Research in Organizational Behavior**, 12: 295–336.
- Powell, W. W., Koput, K. W., & Smith-Doerr, L. 1996. Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology. **Administrative Science Quarterly**, 41(1): 116–145.
- Prasad, J. 1935. The psychology of rumor: A study relating to the great Indian earthquake of 1934. **British Journal of Psychology**, 26: 1–15.
- Preece, J. 2000. **Online Communities: Designing Usability, Supporting Sociability**. John Wiley & Sons.
- Preece, J., & Maloney-Krichmar, D. 2005. Online Communities: Design, Theory, and Practice. **Journal of Computer Mediated Communication**, 10(4): 00–00.
- Quantcast (2011) Macrumors.com- Web site Audience Profiles from Quantcast. Retrieved November 17, 2011 from <http://www.quantcast.com/macrumors.com>
- Ramage, D., Rosen, E., Chuang, J., Manning, C. D., & McFarland, D. A. 2009. Topic modeling for the social sciences. **NIPS 2009 Workshop on Applications for Topic Models: Text and Beyond**.
- Ramírez, R. 1999. Value Co-Production: Intellectual Origins and Implications for Practice and Research. **Strategic Management Journal**, 20(1): 49–65.
- Rao, H. 2002. Interorganisational ecology. In J. C. Baum (Ed.), **Companion to organizations**: London: Blackwell.
- Reger, R. K., & Huff, A. S. 1993. Strategic groups: A cognitive perspective. **Strategic Management Journal**, 14(2): 103–123.

- Rheingold, H. 2000. **The Virtual Community: Homesteading on the Electronic Frontier** (revised edition.). The MIT Press.
- Rindova, V. P., & Fombrun, C. J. 2001. Entrepreneurial action in the creation of the specialty coffee niche. In C. B. Schoonhoven & E. Romanelli (Eds.), **The entrepreneurship dynamic**. Stanford: Stanford University Press.
- Rindova, V. P., & Petkova, A. 2007. When Is a New Thing a Good Thing? Technological Change, Product Form Design, and Perceptions of Value for Product Innovations. **Organization Science**, 18(2): 217.
- Rogers, E. M. 1995. **Diffusion Of Innovations** (4th ed.). Free Press.
- Rosa, J. A., Porac, J., Runser-Spanjol, J., & Saxon, M. 1999. Sociocognitive Dynamics in a Product Market. **The Journal of Marketing**, 63: 64–77.
- Rosch, E. 1975. Cognitive representations of semantic categories. **Journal of Experimental Psychology: General**, 104(3): 192–233.
- Rosch, E., & Mervis, C. B. 1975. Family resemblances: Studies in the internal structure of categories. **Cognitive psychology**, 7(4): 573–605.
- Rose, A. 1951. Rumor in the Stock Market. **Public Opinion Quarterly**, 15(3): 461–486.
- Rosnow, R. L. 1991. Inside rumor: A personal journey. **American Psychologist**, 46(5): 484–496.
- Rosnow, R. L., & Fine, G. A. 1976. **Rumor and gossip**. Elsevier.
- Ruef, M. 1999. Social Ontology and the Dynamics of Organisational Forms: Creating Market Actors in the Healthcare Field, 1966–1994. **Social Forces**, 77(4): 1403–1432.
- Sanchez, R., & Mahoney, J. T. 1996. Modularity, Flexibility, and Knowledge Management in Product and Organization Design. **Strategic Management Journal**, 17: 63–76.

- Santos, F., & Eisenhardt, K. 2009. Constructing Markets and Shaping Boundaries: Entrepreneurial Power in Nascent Fields. **Academy of Management Journal**, 52(4): 643–671.
- Saussure, F. de. 1966. **Course in General Linguistics**. (W. Baskin, Tran.). McGraw-Hill.
- Saxenian, A. 1994. **Regional Advantage: Culture and Competition in Silicon Valley and Route 128**. Harvard University Press.
- Schau, H. J., Muniz, A. M., Jr, & Arnould, E. J. 2009. How brand community practices create value. **Journal of Marketing**, 73(5): 30–51.
- Scott, W. R. 2000. **Institutions and Organizations**. SAGE Publications, Incorporated.
- Seidel, V. 2005, December. **MANAGING NOVEL PRODUCT CONCEPTS: A PROCESS THEORY**. Stanford University.
- Shah, S. 2000. **Sources and Patterns of Innovation in an Consumer Products Field: Innovations in Sporting Equipment**, Sloan School of Management, Massachusetts Institute of Technology, Cambridge.
- Shah, S. K., & Tripsas, M. 2007. The accidental entrepreneur: the emergent and collective process of user entrepreneurship. **Strategic Entrepreneurship Journal**, 1(1-2): 123–140.
- Shapin, S. 1996. **The scientific revolution**. University Of Chicago Press.
- Shaw, A., & Benkler, Y. 2012. A Tale of Two Blogospheres. **American Behavioral Scientist**, 56(4): 459–487.
- Shibutani, T. 1966. **Improvised News: A Sociological Study of Rumor** (First Edition, Second Impression.). Irvington Pub.
- Shibutani, T. 1986. **Social processes**. Univ of California Press.
- Schilling, M. A. 2005. A “Small-World” Network Model of Cognitive Insight. **Creativity Research Journal**, 17(2-3): 131–154.

- Shwed, U., & Bearman, P. S. 2010. The Temporal Structure of Scientific Consensus Formation. **American Sociological Review**, 75(6): 817–840.
- Simakova, E., & Neyland, D. 2008. Marketing mobile futures: assembling constituencies and creating compelling stories for an emerging technology. **Marketing Theory**, 8(1): 91–116.
- Snow, D. A. 2004. Framing processes, ideology, and discursive fields. **The Blackwell companion to social movements**.
- Spencer, J. W. 2003. Global Gatekeeping, Representation, and Network Structure: A Longitudinal Analysis of Regional and Global Knowledge-Diffusion Networks. **Journal of International Business Studies**, 34(5): 428–442.
- Sproull, L., Dutton, W., & Kiesler, S. 2007. Introduction to the Special Issue: Online Communities. **Organization Studies**, 28(3): 277–281.
- Star, S. L., & Griesemer, J. R. 1989. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. **Social Studies of Science**, 19(3): 387–420.
- Stelter, B. 2008. My Son, the Blogger: An M.D. Trades Medicine for Apple Rumors. **New York Times**.
- Stone, B. 2009. Would You Buy the Mythical Apple Tablet? **New York Times**. <http://bits.blogs.nytimes.com/2009/07/28/would-you-buy-the-mythological-apple-tablet/>.
- Stinchcombe, A. L. 1990. **Information and Organizations**. University of Calif. Press.
- Swanson, E. B., & Ramiller, N. 1997. The Organizing Vision in Information Systems Innovation. **Organization Science**, 8(5): 458–474.
- Sunstein, C. R. 2009. ***On rumors: how falsehoods spread, why we believe them, what can be done***.

- Tarde, G. 1898/1969. The public and the crowd. **On Communication and Social Influence. Selected Papers edited and with an Introduction by Terry N. Clark**, Chicago, 277–294.
- Thomke, S., & Hippel, von, E. 2002. Customers as innovators. **Harvard Business Review**.
- Tripsas, M. 1997. Unraveling the process of creative destruction: Complementary assets and incumbent survival in the typesetter industry. **Strategic Management Journal**, 18(s 1): 119–142.
- Tripsas, M. 2008. Customer preference discontinuities: a trigger for radical technological change. **Managerial and Decision Economics**, 29: 79–97.
- Tsoukas, H. 2009. A Dialogical Approach to the Creation of New Knowledge in Organizations. **Organization Science**, 20(6): 941–957.
- Tsui, C., Wang, P., Fleischmann, K., & Oard, D. W. 2009. Understanding it innovations through computational analysis of discourse. **ICIS 2009 Proceedings**.
- Tushman, M. L., & Anderson, P. 1986. Technological Discontinuities and Organisational Environments. **Administrative Science Quarterly**, 31(3): 439–465.
- Utterback, J. M. 1996. **Mastering the dynamics of innovation**. Harvard Business School Press
- Utterback, J. M., & Abernathy, W. J. 1975. A dynamic model of process and product innovation. **Omega**, 3(6): 639–656.
- Utterback, J. M., & Suárez, F. F. 1993. Innovation, competition, and industry structure. **Research Policy**, 22(1): 1–21.
- Vaast, E. 2007. What Goes Online Comes Offline: Knowledge Management System Use in a Soft Bureaucracy. **Organization Studies**, 28(3): 283–306.
- Van de Ven, A. H. 1993. A community perspective on the emergence of innovations. **Journal of Engineering and Technology Management**, 10(1-2): 23–51.

- van Ginneken, J. 2003. **Collective Behavior and Public Opinion: Rapid Shifts in Opinion and Communication** (European Institute for the Media Series). Routledge.
- Van House, N. 2003. Science and Technology Studies and Information Studies. **Annual Review of Information Science and Technology**, 1–85.
- Vance, A. 2010. Microsoft and H.P. to Reveal Slate PC Ahead of Apple. **New York Times**. <http://bits.blogs.nytimes.com/2010/01/05/ahead-of-apple-microsoft-and-hp-to-reveal-slate-pc/>.
- Wang, P. 2009. Popular concepts beyond organizations: Exploring new dimensions of information technology innovations. **Journal of the Association for Information Systems**, 10(1): 1–30.
- Wang, P., & Ramiller, N. 2009. Community learning in information technology innovation. **MIS quarterly**, 33(4): 709–734.
- Watts, D. J. 2004. The “New” Science of Networks. **Annual Reviews in Sociology**, 30: 243–270.
- Watts, D. J. 2004. **Six Degrees: The Science of a Connected Age**. W. W. Norton.
- Watts, D. J. 2007. A twenty-first century science. **Nature**, 445(7127): 489–489.
- Weber, K., & Dacin, M. T. 2012. The cultural construction of organisational life: Introduction to the special issue. **Organization Science**.
- Weber, K., & Glynn, M. A. 2006. Making Sense with Institutions: Context, Thought and Action in Karl Weick’s Theory. **Organization Studies**, 27(11): 1639–1660.
- Weick, K. 1995. **Sensemaking in Organizations (Foundations for Organisational Science)**. Sage Publications, Inc.
- Wellman, B., & Gulia, M. 1999. Net surfers don't ride alone: Virtual communities as communities. **Networks in the global village**.

- Welser, H. T., Gleave, E., Fisher, D., & Smith, M. 2007. Visualizing the signatures of social roles in online discussion groups. **Journal of Social Structure**, 8(2): 564–586.
- West, J., & Lakhani, K. R. 2008. Getting Clear About Communities in Open Innovation. **Industry & Innovation**, 15(2): 223.
- White, H. C. 1981. Where Do Markets Come From? **American Journal of Sociology**, 87(3): 517–547.
- White, H. C. 1992. **Identity and Control: A Structural Theory of Social Action**. Princeton Univ Press.
- White, H. C. 1995. Network Switchings and Bayesian Forks: Reconstructing the Social and Behavioral Sciences. **Social Research**, 62(4): 1035–1063.
- White, H. C. 2000. Modeling discourse in and around markets. **Poetics**, 27(2-3): 117–133.
- White, H. C. 2004. **Markets from Networks: Socioeconomic Models of Production**. Princeton University Press.
- Williamson, O. E. 1999. Strategy Research: Governance and Competence Perspectives. **Strategic Management Journal**, 20(12): 1087–1108.
- Wittgenstein, L. 1958. **Philosophical Investigations**. Basil Blackwell.
- Wry, T., Lounsbury, M., & Glynn, M. A. 2011. Legitimizing Nascent Collective Identities: Coordinating Cultural Entrepreneurship. **Organization Science**, 22(2): 449–463.
- Xie, Z., Nelson, P.C., Xiao, W., & Tirpak, T. M. 2004. Using Noun Phrase Centrality to Identify Topics for Extraction based Summaries. **Proceedings of the IASTED International Conference Knowledge Sharing and Collaborative Engineering**.
- Zerubavel, E. 1999. **Social mindscapes: An invitation to cognitive sociology**.
- Zheng, H.-T., Kang, B.-Y., & Kim, H.-G. 2009. Exploiting noun phrases and semantic relationships for text document clustering. **Information Sciences**, 179(13): 2249–2262.

- Zuckerman, E. W. 1999. The Categorical Imperative: Securities Analysts and the Illegitimacy Discount. **American Journal of Sociology**, 104(5): 1398–1397.
- Zukin, S., & DiMaggio, P. 1990a. Introduction. In S. Zukin & P. DiMaggio (Eds.), **Structures of Capital: The Social Organization of the Economy**: 1–36. Cambridge, UK and New York: Cambridge University Press.
- Zukin, S., & DiMaggio, P. (Eds.). 1990b. **Structures of Capital: The Social Organization of the Economy**. Cambridge University Press.

Appendices

Annexe 1: Data collection methods

Crawling was performed using custom PHP scripting with the cURL utility. Initial URLs were inputted and pages then analysed for patterns of URL formation. In line with other vBulletin installations, a pattern for the URL formation for threads was in the form of *http://forums.macrumors.com/showthread.php?t=1407748*, where the number corresponded to a unique id for the thread. Retrieving this base URL resulted in the first page of a thread, showing the initial 25 posts. If more posts were present in the thread, then a box was present indicating how many total pages there were. This first page was processed using pattern matching and the range of pages for the thread were determined. Subsequent pages in the thread were accessed in the following general pattern *http://forums.macrumors.com/showthread.php?t=1407748&page=2*, *"/showthread.php?t=1407748&page=3*, ... *"/showthread.php?t=1407748&page=n*. Each page used a templating engine to display posts in an identical manner. Posts were then extracted by analysing the HTML source-code looking for the template pattern.

Extraction of the data was conducted by a crawler script that was set to repeat on a 1 minute time-out on a remote server. This was in order to not flood the site with traffic and to provide minimal load on the server. Total time to crawl the site took 2.5 days, and this resulted in 15GB worth of textual data. This was then processed locally and redundant elements were removed. The resulting text corpus was roughly 10GB.

The extracted data was stored in a relational database, set up in MySQL. From the MySQL, data analysis was performed through several interfaces, including PHP and Python scripting. The latter interface allowed for increased flexibility and access to text analysis frameworks

such as the HTML parser BeautifulSoup. Such a tool afforded the ability to deal with raw HTML code as if it were simply text. An advantage of building a interface layer on top of a database was efficient data retrieval and multiple modes of data exploration. It should be noted that standard word processing tools such as Microsoft Word cannot handle such large amounts of data.

Annexe 2: Alphabetical list of part-of-speech tags used in the Penn Treebank Project

Source: http://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

Number	Tag	Description
1	CC	Coordinating conjunction
2	CD	Cardinal number
3	DT	Determiner
4	EX	Existential <i>there</i>
5	FW	Foreign word
6	IN	Preposition or subordinating conjunction
7	JJ	Adjective
8	JJR	Adjective, comparative
9	JJS	Adjective, superlative
10	LS	List item marker
11	MD	Modal
12	NN	Noun, singular or mass
13	NNS	Noun, plural
14	NNP	Proper noun, singular
15	NNPS	Proper noun, plural
16	PDT	Predeterminer
17	POS	Possessive ending
18	PRP	Personal pronoun
19	PRP\$	Possessive pronoun
20	RB	Adverb
21	RBR	Adverb, comparative
22	RBS	Adverb, superlative
23	RP	Particle
24	SYM	Symbol
25	TO	<i>to</i>
26	UH	Interjection
27	VB	Verb, base form
28	VBD	Verb, past tense
29	VBG	Verb, gerund or present participle
30	VBN	Verb, past participle
31	VBP	Verb, non-3rd person singular present
32	VBZ	Verb, 3rd person singular present
33	WDT	Wh-determiner
34	WP	Wh-pronoun
35	WP\$	Possessive wh-pronoun
36	WRB	Wh-adverb

Annexe 3: Advanced rules for method in collapsing and labelling nodes

- 1) Find the head word for each Noun Phrase Chunk, use this as a node label placeholder, saving full Noun Phrase in a list and keeping track of the frequency of occurrence in the text. As each new noun phrase is added to the mass list of nodes, a string similarity comparison is done with the DICE algorithm to determine an existing match within the appropriate threshold
- 2) If noun phrases appear in the same sentence, an edge is formed between them;
- 3) a) Each Node is analysed according to its Noun Phrase list. Next, the most frequently occurring Noun Phrase in the list is analysed. If this NP accounts for 20 percent of the sum frequency for the node and the string difference is below the threshold, it is substituted as the Node Label
 b) the second most frequent Noun Phrase in the list for the node is analysed. If this NP accounts for 20 percent of the sum frequency for the node and the string difference is below the threshold, it is appending with a “ / ” to the Node Label; 3c) if none of the noun phrases account for 20 percent of the sum node frequency, then a “*” is appended to the node label and it is left at the head word.
- 4) Traverse the network and remove any nodes that have less than three connections. This pruning cleans up the network and improves software analysis performance
- 5) Between-centrality (Freeman, 1977) is calculated for nodes in the network.

Annexe 4: Descriptions of tech-blogs from an aggregator

Site descriptions gathered from site aggregator Quantcast.com

Macrumors

“MacRumors.com is a website focused on Apple news and rumors. As the most popular Mac website, the site attracts a broad audience of both consumers and professionals interested in the latest technologies and products. The site also boasts an active community focused on purchasing decisions and technical aspects of the iPod and Macintosh platforms. With breaking news, live event coverage, discussion forums, tutorials and a buyer's guide, *MacRumors.com commands the attention of individual customers and industry watchers alike.*”

Engadget

“Technology gadget tips and reviews.”

BoyGeniusReport

“BGR.com is a leading online destination for news and commentary focused on the mobile and general consumer electronics markets. It is the number one source of exclusive and breaking mobile news, and a technology category leader among early adopters, savvy technophiles and casual readers alike.”

Crackberry

“A site dedicated to CrackBerry users and abusers featuring news coverage, how to guides, forums and BlackBerry software, themes and accessories. This site reaches over 2.7 million monthly people, of which 690K (26%) are in the U.S. The typical visitor owns a Blackberry and reads CNET.com.”