

Organizational identity and competition

a study of US Semiconductor industry

Arttu Säynäjäkangas

Master's thesis

Sociology

Department of Philosophy and Social Sciences (YFI)

Autumn 2017

Tiivistelmä

Organisationaalinen identiteetti ja kilpailu

Tutkielma puolijohdeteollisuudesta

Arttu Säynäjäkangas

Sosiologia, yhteiskuntatieteiden maisterintutkinto-ohjelma

Pro gradu-tutkielma

Yhteiskuntatieteiden ja filosofian laitos

Jyväskylän yliopisto

Ohjaaja: Markku Lonkila

Syksy 2017

sivumäärä: 52 sivua + liitteet 13 sivua

Tutkimuksen tehtävänä on selvittää, miten kaksi samalla puolijohdesektorilla kilpailevaa yritystä, Intel Corporation ja Advanced Micro Devices, kuvailee itseänsä ja kenttäänsä suhteessa kilpailijoihinsa ja siten rakentaa ja ylläpitää organisationaalista identiteettiään. Tutkimuksen teoreettisena näkökulmana on tarkastella markkinoita sosiaalisesti rakentuneena kenttänä jota kilpailu ylläpitää. Analyysimenetelmänä on käytetty koneoppivaa tekstin mallinnusmenetelmää nimeltä Latentti Dirichlet Allokaatio (LDA), jolla on tutkittu tutkimukseen valittujen yritysten vuosiraportteja. Menetelmällä tarkastellaan vuosiraporteissa esiintyviä aiheita sekä niiden vuosittaista muutosta. Havainnot tarkennetaan suorilla lainauksilla teksteistä. Tulosten pohjalta argumentoin muun muassa, että tutkimuksen yrityksistä pienempi tarvitsee organisationaalisen identiteettinsä ylläpitämiseksi jatkuvaa peilausta isompaan kilpailijaansa kun taas markkinalla hallitsevassa asemassa oleva isompi yritys näkee itsensä yhtenä toimijana muiden joukossa. Tutkimuksen lopuksi esitän kriittisiä havainnot menetelmän käytöstä tutkimuksessa, sekä mahdollisia tulevaisuuden tutkimuksellisia polkuja, joita tutkielma osaltaan avasi.

Tutkielman avainsanat: organisationaalinen identiteetti, markkinoiden sosiologia, LDA, puolijohdeteollisuus

Table of Contents

Tiivistelmä	1
0 - Abstract	4
1 - Introduction	5
1.1 - The aim of this study	5
1.2 - The semiconductor industry	6
1.2.1 - Semiconductor Industry	6
1.2.2 - AMD & Intel	8
1.3 - The structure of this thesis	9
2 - Theoretical framework	11
2.1 - The big picture	11
2.1.1 - Sociology of markets	11
2.1.2 - A sociological theory of action in the market	13
2.1.3 - Conception of control	16
2.2 - Organizational identity and competition	18
2.2.1 - Organizational identity	18
2.2.2 - Strategic Group Identity	20
2.3 - Conclusion	22
3 - Data and Methods	24
3.1 - Data	24
3.1.1 - Annual Reports as Data	24
3.1.2 - Form 10-K	26
3.2 - Method	27
3.2.1 - Big Data in Social Sciences	27
3.2.2 - Topic Modeling	29
3.2.3 - Latent Dirichlet Allocation	30
3.3 - Use of the Data and Method in this study	33
4 - Results	36
4.1 - Foreword to results	36
4.2 - AMD	36
4.2.1 - Intel as aggressive competitor -topic	37
4.2.2 - Intel and component market -topic	39
	2

4.2.3 - Graphic chipset market -topic	41
4.3 - Intel	42
4.3.1 - Product segments and market -topic	43
4.3.2 - Industry standards -topic	45
5 - Discussion and conclusion	47
5.1 - Review of findings	47
5.2 - Discussion on methodology	48
5.3 - Conclusion	51
6 - References and Appendix	53
6.1 References	53
6.2 - Appendix	61
6.2.1 - Full LDA model outputs	61
AMD	61
Intel	63
6.2.2 - The code for LDA	64
6.2.3 - Links to the sec database	66

0 - Abstract

This thesis investigates how firms construct their organizational identity through competition with other firms that they deem similar through a close study of two notable semiconductor companies. The annual reports that the firms produce on a yearly basis is a promising cache of data into understanding social phenomena that happen in and between firms, including those relating to organizational identity. A relatively novel machine learning technique called Latent Dirichlet Allocation is used to model this data in the case of two prominent semiconductor firms, Advanced Micro Devices and Intel corporation, to inductively discover some aspects of how they discuss competition in annual reports which is then used to deepen our understanding of competition as a facilitator of organizational identity.

1 - Introduction

1.1 - The aim of this study

“Should we be in the MP3 business? How does that fit with who we are”
Board member of Linco, a photography company (Tripsas 2009)

“Who are we as a company” is a question that is often asked in board meetings of companies and even by ordinary employees having a small talk by the water cooler at their workplace. The answer to this question guides the decision making of firms, helps maintain solidarity and sense of purpose for the employees of the company and also helps its potential clients and customers to understand how the company can serve their specific needs. This question of identity of an organization is one that has gained increasing attention in recent past in organizational studies (Gioia 2008). One reason for this might be that identity in general has become an increasingly salient issue in contemporary postmodern lives. As we experience an ever faster pace of change in our society, we look for something tangible that helps us make sense and grasp our social reality. The purpose of this thesis is to approach the concept of organizational identity from a distinctly sociological point of view by looking at how it is constructed in social interaction. For private businesses, one of the main component of social interaction happens with other similar firms, that is, the *competitors*. In this paper I will look into how two prominent firms in the semiconductor industry discuss and construct their identity by analyzing how they discuss competition.

Competition in markets is the great organizing principle of private enterprise in modern capitalist societies. However competition does not necessarily happen on its own but often needs help from the state. One of the major ways the state can help in maintaining healthy competition between firms is through antitrust legislation. If a corporation grows too big, it will have ability to control competition, for example by buying off smaller rivals. In this paper I focus on one major industry, the semiconductor industry and on two major corporations that have a stake in the market, to further our understanding of how competition structures markets. The main objective of this study is to make a scientific inquiry into how

two firms, operating on a same market, perceive their competitive environment and what kind of differences they have in their perception on this matter. This will then be used to further our understanding of organizational identity in relation to competition. First I will briefly introduce the semiconductor industry and the two corporations and then move on to explain the structure of this thesis.

1.2 - The semiconductor industry

1.2.1 - Semiconductor Industry

The birth of the semiconductor industry is generally dated to the 1960s when the technology behind modern computers became commercially viable. Today, a device with a semiconductor functions by exploiting the electronic properties of semiconductor materials (most commonly silicon) to process information, typically in an integrated circuit, such as the microprocessor found in personal computers. The global semiconductor industry has been one of the drivers of information technology revolution which has seen global economy being integrated through the power of the microchip that enables computers to process information with ever increasing velocities. The semiconductor industry today is dominated by several large multinational corporations operating mainly from USA, Europe and Asia. There are numerous reasons to why I chose firms operating in the semiconductor industry for this study. One such reason is because they have been subject to a wide range of studies in organizational research. They also serve in the vanguard of how technological advancements in information technology made possible the long period of growth in the global economy from 1950s onwards.

In this thesis I will study a particular aspect of organizational identity of two notable semiconductor firms, Intel and Advanced Micro Devices (AMD). There are multiple reasons to why I chose these two companies. Perhaps the of the most prominent one is the fact that these two corporations have been competing with each others for decades. Especially for businesses operating in information technology sector, such stability in a sector's competitive structure is intriguing and should make an interesting research subject.

Macher et al. (2007) studied the global semiconductor industry and established that the sector was undergoing several forms of structural changes in the 2000s. Arguably the most significant of those was the perceived shift from “integrated device manufacturers” (whereby the companies both design and manufacture hardware), to more vertically specialized firms that focus either on design and marketing (“fabless”) or manufacturing (“foundries”). Jiang et al. (2011) investigated how incumbent firms in semiconductor industry approach innovation. They assert that much of the literature in the field is focused on explaining why and how dominant firms are fail to adapt to changing technological and organizational paradigms. Incumbents are generally presumed to focus on incremental innovations on their dominant products in the market and tend to neglect emerging and potentially disrupting innovations. Jiang et al. however find that incumbents in the semiconductor sector invest considerably in finding alternative and superior inventions to their dominant products.

The semiconductor sector is relatively stable industry with exits and entrances by old and new businesses fairly uncommon. A major reason for this is the high cost of designing and manufacturing of semiconductor products. A single microchip today can hold millions of transistors which are produced in a delicate and slow process. As I already argued, this relative stability of semiconductor industry makes it a suitable target for research in social science when investigating temporal changes in their corporate culture, which in this study is the organizational identity.

In this study I focus on two of the most prominent semiconductor companies, Advanced Micro Devices (AMD) and Intel Corporation (Intel). To study how these two companies discuss competition. I argue that the way a company, which in this study means its top decision making bodies, discusses competition provides us with some insights into how it perceives itself and its place in the economy and even the wider society.

1.2.2 - AMD & Intel

AMD and Intel are the most prominent players in the consumer semiconductor business. They have been long standing rivals in the market for microprocessor chips and integrated circuits with such well-known product brands as Pentium and Athlon competing over the personal computer CPU (Central Processing Unit, the microchip) market. Intel was incorporated in July of 1968 and AMD next year in May. In 1981, the largest firm in the IT-sector, IBM entered the personal computer market and wanted Intel to manufacture its microchips on the condition that Intel provides a second-source manufacturer for its patented x86 microprocessors for which AMD was chosen. This requirement was made by IBM to make sure they wouldn't grow too dependent on one supplier for their components. (Singer 2012). Intel and AMD have shared close bonds ever since.

While the two companies are fairly similar in terms of customer base, product portfolio and global interests, there have been some important actions that have taken the two companies on different paths during the 2000s. While AMD had strongly pivoted towards fabless strategy whereby it divested its global manufacturing to focus on designing and marketing processes, Intel still has several production facilities all over the globe in 2017.

In general, AMD has always been considered the underdog between the two (Kobylecky 2016). Intel has been keen to use its dominant position to control its weaker rival, often using methods that have been deemed illegal in many different jurisdictions and for which Intel has been forced to pay fines and settlements. AMD has often claimed that Intel has a monopoly position in the market for microprocessors which would give it certain responsibilities to maintain competitive environment due to antitrust laws that seek to ensure competitive markets by forbidding any single corporation to control too much of it. For example in 2009 Intel was forced to pay a fine of over 1 billion euros by the EU for abusing its dominant position in the x86 architecture processor market. One of the illegal actions Intel had taken was paying Original Equipment Manufacturers (OEMs) not to use rival products, which would have been most commonly supplied by AMD. AMD's position regarding competition in the microprocessor market can be seen from its then CEO Hector Ruiz's keynote made for

American Antitrust Association in 2007: *“There is no proper or defensible place for illegal monopolies in the 21st century global marketplace...My purpose is not to argue for competitive advantage - we know how to compete. My purpose is to lay out the facts so that law and economics can do their job to protect consumers”* (BetaNews 2007). AMD has been keen to argue its place in the market as the protector of consumer choice. If Intel’s market share grows too large, the consumers will have less choice and there will be less competition and therefore innovation in the sector.

In 2011, AMD spokesman Michael Silverman said in an interview that AMD would start focusing less on Intel in their business strategy: *“Any time a company is more focused on its competitor than what the market wants and needs, it is probably not in a good place for long-term growth”*, continuing: *“Our CEO Rory Read is driving a shift internally to think beyond the typical and outdated AMD/Intel mindset.”* (Gaudin 2011). However as some analysts note, it’s hard to simply push away from such an important business relationship as what AMD and Intel have had for so long. The point of interest about this dynamic of competition between these two firms for this study is how it structures the market and the firm’s understanding of itself in it through the constant negotiation of what is a fair and competitive relationship between the two firms.

Next I will discuss how this thesis is structured and present the research question which I seek to answer. As I delve into the dynamics of this relationship between AMD and Intel. I will do so by focusing on a repository of public data issued by both companies between 2004-2016. These documents are called annual reports which I will introduce in chapter 3.1.

1.3 - The structure of this thesis

In the next part I will start from general theoretical discussions pertaining to this study in the field of economic sociology. First I will elaborate my approach to the firm and the market in modern capitalist society (part 2.1), and how organizational identity works as a useful conceptual tool to understand some actions firms take in the market (part 2.2) and how such

action might be visible in the corporate public disclosures which I have chosen to use as the data of this study. In the part 3.1 I will look into how these public documents, commonly called annual reports, have been used before as data in social sciences. In part 3.2 I will introduce the method used in this study which is applied to simplify this data in a manner that makes it easier for us to find out larger trends through the observation periods. In chapter 4 I will display the results obtained from combining the data and method and discuss how it can further our understanding of organizational identity. In chapter 5 I conclude by first discussing both advantages and deficiencies of using this combination of data and methods in creating new understanding on how the firms discuss their competitive environment, how it affects firms actions in markets and ultimately their organizational identity. Finally I discuss what kind of possible future paths this could be taken to research from this study.

2 - Theoretical framework

2.1 - The big picture

2.1.1 - Sociology of markets

Before I delve into the theory of organizational identity, I will introduce the general framework of this thesis. I will present the private business, or firm, in a capitalist system as an organization operating in one or several “markets”. The primary objective of the firm is to produce profit which is then divided between the stakeholders, which for publicly traded companies can be divided into employees, managers and shareholders. As the general area of my thesis, the capitalist market system, is commonly regarded as the domain of economics, I will first discuss the general differences between economics and sociology and then elaborate on some relevant concepts and theories pertaining to this study. There are many sociological theories and approaches to understand to help understand how and why firms operate in their markets the way they do. While the main focus of this study is on organizational identity, I will introduce some other sociologically relevant theories that may aid in understanding the central theoretical concept of this thesis.

Economic sociology, as defined by Fligstein is the “*study of how the material production and consumption of human populations depend on social processes for their structure and dynamics*” (Fligstein 2001, p. 6). Fligstein approaches markets as the production side of a duality where people both produce and consume goods in modern world. Furthermore he argues that economic sociology needs a solid organizing frame on how to understand markets, which he claims should stand on an institutional grounding. The purpose of institutions is to create stability in human interaction and social life (Berger & Luckmann 1991). Fligstein then approaches markets as what sociologists call fields consisting of sets of institutions build to promote the survival of participating or key firms. Fligstein, following White (1981), define market as “*a self-reproducing role structure of producers*”. One of the most distinguishing features of a market society, as Fligstein notes, is the fact that new technologies and players can enter a given market, forcing the established players to adapt

and constantly look out for possible disturbances to the existing role structure. Fligstein (2001) presents his approach, which he calls the *political-cultural*, as a general framework that can better explain many sociologically curious phenomena occurring in the market system. As the name “political-cultural” suggests, Fligstein’s approach attempts to bring questions of politics and power into the debate of how a market system functions. These key questions are, as proposed by Fligstein, how markets are initially created, who gets to be a player on the field, what are the rules on the field and who defines them. This approach replaces the profit-maximizing actor with people who are simply trying to promote the survival of their firm. (Fligstein 2001, p. 10-17.) Fligstein posits that a field typically consists of incumbents, the actors who are able to reap the benefits from the rules of the field, and of challengers, who try to challenge the prevailing rules and conceptions and to alter them in the challenger’s benefit.

Podolny et al. in their paper ‘*Networks, Knowledge, and Niches: Competition in the Worldwide Semiconductor Industry, 1984-1991*’ (1996) studied and developed the concept of market niche by exploring the global semiconductor business. The paper draws a lot of inspiration from population ecology and network studies. The authors start off by dispensing with the idea that ideas and inventions arise spontaneously and in isolation from a firm’s investment in research and development but are instead produced through firm’s cooperation with other producers of knowledge. Podolny et al. proposed that firms in a market have to deal with two forces that will affect their life-chances: crowding and status. When a new market is created, it will initially have less actors and therefore less competition, in other words it’s less crowded. When competition tightens and more competitors emerge to fight for survival in the market, the status of the firms becomes an important part of survival. The higher the company’s status is, the better its life-chances are. In the context of this study, Intel can be considered a high-status player which has been able to sustain its position, not only due to its knowledge and know-how to building high quality products, but also due to its status as the semiconductor manufacturer. Other firms, such as AMD need to keep Intel on its eyes if they are to stay relevant in the field.

One way to understand the sociological view to what markets are can be found from Ahrne et al. (2015) who propose that markets and organizations (such as firms) are not as separate, or

even opposite, as is commonly understood. They posit that markets themselves can be understood as organizations, especially in a modern market society, and point out some of the many organizing features a market typically has: rules, sanctions, monitoring of the members and so on. Markets as organization is, as the scholars propose, a conceptual tool to better help grasp the somewhat mysterious nature of the market.

2.1.2 - A sociological theory of action in the market

The theoretical cornerstone of this thesis is the sociological approach to action in markets that emphasizes the social embeddedness of all economic activity. The first to formulate a general framework in this field was Karl Polanyi who used embeddedness to describe how the economy is embedded in society's social structure (Polanyi 1957). This view was more recently developed by Granovetter in his widely cited paper *Economic Action and Social Structure: The Problem of Embeddedness* (1985). Granovetter argues that all economic action is deeply social and, as he called, "embedded" in the social structure of society. Neil Fligstein then further elaborated on this viewpoint to social action in markets in his book *The Architecture of Markets* (2001). Fligstein's political-cultural approach starts with the idea that markets can be thought of as what sociologists call "fields". A field is an organized social space which contains actors who attempt to create a system of domination in the space. Key question in this approach is "*what social rules must exist for markets to function, and what types of social structures are necessary to produce stable markets*"? In a book *A Theory of Fields* (2012), Fligstein and McAdam further develop the microfoundations of how they approach social action on fields. I take this as a further development to Fligstein's earlier work in *The Architecture of Markets* to how markets can be understood from a sociological point of view. I will discuss this theory in more detail later, but first I will how the capitalist market system is understood in this thesis. It starts off by investigating the differences between how sociology and economics understand a capitalist economy.

The debates around the question concerning division of labour between sociology and economics has been substantial. One way to create distinctions between the two sciences is to look at how they approach human action. Much of contemporary economics is based on the theory of rational action. It starts with the assumption that humans are capable, if not have

inherent tendency towards rational, utility maximizing action. In other words, it makes certain assumptions about humans and their nature, handily encapsulated in the term *homo oeconomicus*. As Fligstein & McAdam (2012, p. 34) note, this theoretical starting point is rather marginal in sociology, and most sociologists are even downright hostile to the theory. Joas (1996), after revisiting past hundred years of sociological theory building on the subject, makes his own effort on creating a refined basis for sociology as a branch of science based on a well-founded and distinctly sociological theory of social action. He lays out a foundation for sociological understanding of human sociability and especially human action which departs from the assumption that human action could be effectively or totally operationalized utilizing the assumption that human action should be understood as behavior that simply attempts to maximize utility, instead presenting his theory of creativity of action.

Fligstein and McAdam, in their book *Theory of Fields* (2012) make a very similar claim, arguing that humans have both a need for material reproduction (the basis for rational models) and for the creation of meaningful explanations in their lives, which is what they call the “*existential function of the social*”. This latter part is which many early sociologists from Emile Durkheim and Max Weber to G. H. Mead have studied in depth (Joas 1996). The theories of Mead are particularly prominent in both Fligstein & McAdams and Joas’ works. One of the central tenets in Mead’s works is the position that individual minds can’t exist in isolation but are a product of human sociability. His theories, that later came to form the school of symbolic interactionism, reject the methodological individualism of economics where organizations emerge out of individual interest. Indeed this methodological starting point has been subject of constant controversy even in economics. These include problems explaining why do rationally acting individuals form firms (Coase 1937) or join social movements (Joas 1996, p. 199-209). For Mead, mind arises out of social act of communication and it is this innate human sociability that makes the collective to what Fligstein and McAdam describe as “*existential refuge*” (2012, p. 40).

Fligsten & McAdam argue that, while generally considered as one of the major antagonists to the mainstream neoclassical economics of the day, Marxist theory building also tends to ignore the fact that humans have this innate desire for creating meaningful symbols and explanations for their lives that can’t be reduced to the mere reproduction of their material

being. As Joas notes, Marx, while taking a position of critique towards the mainstream political economics of his age, never really departed from the core assumptions of rational action. In the marxist historical materialist approach, people are guided by their material interests and even complex social constructs such as religion are ultimately considered means to control the masses. In this study I also approach actions of firms through the concept of interest, however I argue that what makes my approach to this distinctly sociological is that I understand different interests that various actors possess to be changing in a creative process through social interaction.

Even more contemporary economic sociologists are deemed to accept the rationalist scheme far too easily (Fligstein 2001, p. 13). Social structures in markets are considered to be efficiency enhancing, and any activities that do not enhance profitability are seen as simply a deviation from this norm. If this rational, utility maximizing model is taken as a theoretical starting point for analyzing social structure in markets, the approach will inevitably lack independent sociological power to explain social phenomena. After all, such explanations are already present in institutional economics (Fligstein 2001, p. 14). Explaining social institutions with the underlying assumptions of individual self-interest, has the danger of rendering economic sociology indistinguishable from institutional economics even if explained using such traditional sociological concepts as networks or norms.

Fligstein and McAdam present the concept of “social skill” as a micro level foundation to understand how and why people create and sustain meso level social institutions (2012, p. 46). This concept makes certain assumptions of what is to be human, centering around our need for meaning as a ground for our being, that can only be attained through social collaboration. This view of human as a *homo sociologicus* is then used to explain various phenomena outside of the individualist framework employed in economics. Social skill is the micro level mechanism that humans possess which help us to create and sustain meaningful frames through which we understand the world and our place in it.

In this study I will aim to understand the modern corporation and its organization level activity in markets from an authentically “sociological” point of view. I present firms as organizations operating in markets where they not only have a purpose of producing profits to

their stakeholders (the “rational action”), but also a need to create meaningful frames through which their particular place in society can be understood (the “creative”). These meaningful frames have been studied in depth through the concept of organizational identity. For further discussion a clarification is needed: when I use the words “organization”, “company” or similar terms, I will refer to the top level management of the firm, and not all those who have a stake in the firm, like employees and owners. Studies have shown that, when compared with other employees in the company, the top management’s beliefs has the greatest correlation to the actual actions that a firm takes (Livengood and Reger 2010).

2.1.3 - Conception of control

The political-cultural approach emphasizes the inter- and intraorganizational power struggle that happen in the market (Fligstein 2001, p. 128). According to this viewpoint, the firms need a conceptual framework to assess which actions to take in the market. This is what Fligstein calls a *conception of control*. It refers to market-specific agreements between actors (such as employees, managers and owners) in firms and between firms in markets. Fligstein describes conceptions of control as both historical and cultural products. Historical in the sense that they are specific to certain industries and societies in certain times and cultural in that they form a “*set of understandings and practices about how things work in a particular market setting*”. (Fligstein 2001, p. 35.) For instance, Fligstein describes how Apple Inc. failed the battle in the personal computer market against Microsoft in the 1990s by failing to assert its conception of control as the dominant in the market. In the 80s Apple was the leading PC manufacturer but the market was relatively small and occupied mainly by hobbyists. Apple failed to leverage their early position in the market, mainly because Microsoft managed to make their operating system Windows into the *de facto* standard operating system by licensing it to personal computer manufacturers. This was against one of Apple’s core philosophies of approaching the hardware (the physical parts of the computer) and software as inseparable. Trying to have the firm’s core technology (for instance, computer software) to become the industry standard became a dominant conception of control in technology sector (Fligstein 2001 p. 72). Microsoft’s success was evident in the near ubiquity of Windows operating system in PCs, especially from the late 90s onwards.

As I stated earlier, markets in capitalist societies can be understood as fields that have players with different resources and capabilities that help them push for their view of what kind of conceptions of control should dominate in a field. The more dominant a firm is, the more dominating its conception of control is. If smaller firms wish to compete and stay relevant in the market, they need to pay attention to the dominant firms which can use their position of power to define for example what consumers want and what kind of standards get popularized. Like I discussed earlier, AMD has a long history of accusing Intel of using its position to manipulate the semiconductor market in a way that inhibits competition.

There has also been a great deal of research into the rise of the shareholder-value paradigm which came to dominate US firms in the 1970s and 1980s (Lazonick & O'sullivan 2000). Fligstein (2001) described how American corporations came to be dominated by finance conception of control in the 1970s. This conception of control asserts that the modern corporation is, first and foremost, a bundle of assets and the primary objective of the firm is to create value for the shareholder. Shareholder-value approach fundamentally changed how major US firms were operated as their main focus was altered from simply being in, for example a commodity business, into creating maximal value for the shareholders. In a more recent study, Jung (2015) looked into how shareholder-value paradigm was connected to growing popularity of downsizing the workforce of the firm, which prior to shareholder-value paradigm was considered a last-ditch effort but has now become a standard procedure to improve cost-efficiency and boost stock price. Shareholder-value paradigm can be thought of as a conception of control which was adopted as a response to the “stagflation” crisis of poor stock performance and high inflation of 1970s which was a threat to the financial capitalist class survival (Fligstein 2001, p. 152; Harvey 2007).

Fligstein, after reviewing the literature on the owner/management debate of who controls the American corporation, state that a manager-controlled firm tends to prefer equity over debt, decrease the ratios of returns to assets and sales, and have relatively low returns on equity and so on, which reflect the “*management’s propensity to retain or reinvest earnings rather than disburse them to shareholders*” (Fligstein 2001, p 131). Annual reports, as a domain where the top management may propose their vision of the company and its direction, should be an

important arena for communication between the conflicting interests of the owners and managers.

I argue that conception of control and organizational identity are deeply connected concepts that operate on different epistemological levels; Conception of control is to organizational identity what power is to identity: The former is the ability to define the latter. When a firm successfully spreads its conception of control across the market, other firms must adopt identities that are aligned with this conception. For example, if a firm manages to make its product the industry standard, other firms need to follow suit and produce products that adhere to this standard, thus changing how they identify themselves as an organization.

I have now presented the firm in a capitalist society as a battleground for advancing the often conflicting interests of various stakeholders in the firm, such as stockholders, management, workforce and the state. All parties will attempt to promote a corporate culture that is aligned with their interest in the firm, whether it be long-term stability, or high stock performance. As firms try to promote a conception of control in the field which most suits their interests, they seek to articulate what the market is and what their place in it is. Much of this corporate culture is embedded in the concept of organizational identity, to which I now delve into.

2.2 - Organizational identity and competition

2.2.1 - Organizational identity

As defined by Hsu and Hannan (2005, following Polos et al. 2002; Hannan et al. 2005) organizational identity consists of social codes, or sets of rules, specifying the features that an organization is expected to possess. The concept was first formalized by Albert and Whetten (1985) as that which is the most *central, distinctive and enduring* in an organization. Ashforth and Mael (1996) further note how critical organizational identity is for motivating and shaping firms' strategies and actions. They propose that identity is a potential bridge for linking macro and micro phenomena, specifically in the organizational context.

As Gioia (2008) notes, identity is a concept that just resonates with how humans deal with their everyday organizational activities. She argues that organizational identity has gained popularity as an object of study, simply because it “*enquiries into the deepest levels of our sensemaking and understanding*”. an individual’s or organization’s identity is a deeply social phenomena that is constantly being constructed in a reflexive process with other similar entities. I argue that organizational identity is not only a mean to achieve economic ends, such as legitimacy, but serves also as a “collective refuge” as discussed by Fligstein and McAdam (2012). When people work in an organization, they constantly create meaningful frames as a central product of their actions, and large portions of this meaning making is captured by the concept of organizational identity.

Ashforth and Mael start their seminal paper *Organizational identity and strategy as a context for the individual* (1992) by investigating Albert and Whetten’s (1985) original definition that proposes organizational identity as the *central, distinctive and enduring* character of an organization. *Centrality* in terms of organizational identity means it is defined by what is the focal part of its mission, whether it be to “*save souls or manufacture high quality tires*” (Ashforth and Mael 1992). *Distinctiveness* implicates comparisons and relationality with other organizations. If a firm is described to have a specific feature, this feature should be understood as a quality that distinguishes it from its competitors. For example, if a firm is described “aggressive”, it would indicate first and foremost that it is “aggressive” when compared to other firms in the business it operates in, who are likewise more “passive”. The last of the three defining features of organizational identity, *endurance*, is rather self-explanatory, indicating the continuous nature of identity. I would argue that it is also the least important of the three, as even if a company’s central and distinctive features change rapidly, it’s still defined by them.

Tripsas (2009), in her case study of a digital photography company named Linco, noted how the firm had considerable difficulties adopting new and superior technologies because it didn’t fit with their organizational identity. She argues this has two reasons: firstly because identity serves as a filter which makes decision making bodies of a company ignore information about new technologies that don’t fit with their organizational identity, and

secondly because “ *identity becomes intertwined in the routines, procedures, and beliefs of both organizational and external constituents, explicit efforts to shift identity in order to accommodate identity-challenging technology are difficult.* ” (Tripsas, 2009.)

As I now have established, organizational identity greatly affects the strategic decision making that takes place in a company as its management constantly constructs meaning to what the organization is about and how it should operate in the market. Another crucial aspect to consider, and one that I will specifically focus on later in this thesis, is how organizational identity is shaped through competition. In the context of for-profit firms, competition is a question that goes into the heart of the question of identity as it defines the firm’s capability to extract profits and operate. Evans (2015) explored how views on organizational competition correspond with organizational identity discourse as she attempted to bridge the gap between organizational strategy and identity. In her study, she interviewed employees and managers of public radio organizations in the US and found how the concept of distinctiveness in relation to competition and other elements of the firm environment affect companies’ strategic choices. Evan’s posits that distinctiveness, one of the three dimensions of Albert and Whetten’s original definition of organizational identity discussed earlier, was particularly prominent in the interviewees answers regarding how organizational identity was defined.

2.2.2 - Strategic Group Identity

Anand et al. (2012) developed the extensively studied concept from organizational studies called “cognitive strategic group” by mixing it with organizational identity. Strategic group is a concept developed in the organization science, originally referring to a group of firms following a similar strategy (Porter, M. E. 1980). It has a lot of similarities to the concept of “field” in sociology that was presented earlier. Strategic group is a conceptual device which helps the managers of firms to form strategic choices and make decisions on how the firm should operate in its market by grouping together the firm’s competitors. “Strategic group identity” is a further development to this concept that, like organizational identity, is defined by what is central, distinctive and enduring to the identity of the firm in the strategic group.

The “collective consciousness” of the firm management can have multiple strategic groups that affect their strategic choices.

Peteraf and Shanley (1997; following Hatten & Hatten 1987) note that there is an issue regarding the level of analysis when analysing organizations with concepts originally developed to study individual humans, such as identity. Simply put, there is a danger to anthropomorphize organizations. While for a small firm with single decision maker, the cognitive processes of the firm are same as those of its CEO, the larger firm has a group of top level decision makers, making these cognitive processes much more vague. I argue that the annual reports propose a valuable avenue to explore this collective mind of modern corporations and the processes involved with decision making for them. Peteraf and Shanley state that the concept of cognitive strategic group only makes sense if it guides organizations’ decision making. If it is simply an “end product” of categorization process, the concept would be useless.

Starting from Peteraf and Shanley’s work on combining strategic group literature with the concept of organizational identity, Anand et al. (2012) present a new theoretical apparatus to approach firm strategic management which they call “*strategic group configuration*”. It seeks to address some issues pertaining prior literature on strategic groups, including a more refined handling of the manner of understanding how firms choose their strategic groups. They also propose that the concept of strategic group configurations potentially has a better grasp on how firms can associate with multiple strategic groups. The authors claim that firms’ connection with a strategic group relies significantly on its organizational identity.

Anand et al. (2012), following Whetten and Mackey’s (2002) formulations about organizational identity, argue that: “*an organization’s identity can determine both who the firm believes it competes with and who it dismisses as a competitor.*” This position implies there is a measure of arbitrariness as to whom a company considers its competitor. Whetten and Mackey (2002) assert that, as firms face massive amounts of information about their environment and competitors, they rely on their central characteristics to categorize and give order and meaning to this information. This is where organizational identity plays a central role. They further argue that a firm’s organizational identity must specify core characteristics

an organization shares in common with others in the strategic group they are in, and which of them differentiate them. The firms need both, to maintain differences to their competitors to remain competitive, and similarities in order to be considered legitimate (Whetten and Mackey 2002).

2.3 - Conclusion

The defining feature of modern capitalist economy is the use of market system as an organizing device to facilitate competition. If a market is understood as a system of reproducible role-structures between firms as defined earlier in this thesis, then identity should play an integral part in how capitalist market system functions. Organizational identity can be understood as the collection of roles that is attributed to a business in its market relating to its competitors. Different producers often occupy different positions or niches in a product market. For example Apple has positioned itself in the smartphone market as a high-end, high quality phone manufacturer. This position can only be understood if there are other producers that produce cheaper and/or lower quality phones. A system of such niche positions, or roles in a more sociological parlance, is what constitutes a market as it is understood in this thesis. In the case of Intel and AMD in the semiconductor business, we can argue that Intel has occupied the dominant position in the market throughout its history while AMD has positioned itself as the challenger. In the next chapter I will introduce the data and methods with which I will seek to explore the particular dynamics of these two firms as they describe their competitive environment.

As firms compete with one another in a market, they are constantly articulating what are, in their view, good forms of competition and how the firm should be understood in relation to other firms that are close to them in terms of, for example product portfolio or target consumer or client segments. When two firms that operate on a same market, use different language in how they discuss what the market is all about and what the conception of control they try to promote in the market should be, we can witness sociologically interesting phenomena taking place. In economics, a firm is typically seen as a rationally operating unit

which seeks only to maximize profit. While the cultural-historical approach to firms employed in this thesis does take “interest” of various actors in a firm as a micro level starting point for explaining what happens in markets, the sociological viewpoint to action in markets notes that what is defined as “interest” needs creative action by the actors. As Joas (1996) noted, the theories that take utility maximizing as their starting point do not explain where or how organizations determine the utility. For example, Joas notes how Luhmann explains how organizational rationality lies in its capability to be means to those who can determine ends (Joas 1996 p. 150). Determining ends is a creative and social process in which new meaningful frames are constantly being created to understand what the organization is all about.

Market as a field means that what is at stake in the market is constantly being negotiated by the various interest groups that have a stake in the market. Interests themselves change as humans interact with each other and create meaningful frames through which they understand their place in social contexts. As these social processes constantly change the market, the firms seek to adapt to these changes, and through this adaptation their organizational identity changes as well. For example, if a high status player in a market, such as Intel in the semiconductor industry, successfully introduces a new processor architecture standard, its organizational identity changes as well. For example it could articulate being a leading innovator in the market for said products.

In this study I have proposed that organizational identity serves a crucial role in the structure of a market. Similarly to individual identity, it needs to be constantly constructed by the members of a strategic group, which the market consists of. This is a work primarily done by the top management of the firms. A key part in the construction of organizational identity for firms is competition and specifically how the firm perceives it. Different firms that compete in the same business typically have different ideas about what the market is and what their role in that market is. In the next chapter I will propose how I seek to study such processes from the point of view of the firm and specifically its top management.

3 - Data and Methods

3.1 - Data

3.1.1 - Annual Reports as Data

As the focus of this study is on large publicly traded corporations, I will use the words “company”, “firm” and “corporation” interchangeably, referring to the aforementioned type of legal entity. The publicly traded companies have several notable characteristics: for instance, anyone can, in principle, buy shares in the company, receiving a right to a part of the company’s profits and may vote on who will be hired in the top leadership positions in the firm. Public companies also are subject to several laws that may somewhat differ from country to country, however the major features of the legal frameworks tend to be similar throughout the world. For example, most developed countries mandate the publicly traded corporations to file public reports that cover the main parts of their operations. In the United States, governmental body responsible for the administration of these is called Securities and Exchange Commission. These documents, of which the most important are the annual reports, cover mostly the firm’s financial operations, but also typically consist of information about how the firm perceives itself, the markets it operates in and potential risk factors that may affect the company in the future. A comparable but not identical document type is the Letter to Shareholders which is typically signed by the firm’s CEO and which many companies choose to send to the public on a yearly basis. These have been widely popular documents for many scientific studies into corporate culture.

I argue that, as the managers and shareholders of a firm constantly strive to control the firm as proposed by the principal-agent problem, the annual reports present a significant arena where the upper management can promote a vision of the market and the firm’s place in it that is most suitable for them. This is supported by Poole’s (2016) study where he conducted a keyword analysis on the letter to shareholders of two major American banks, Citigroup and Bank of America before, during and after the financial crisis of 2008. Poole discovered that

during the times of crisis, the banks promote messages that assert vision and forward looking strategies of success while distancing the performance issues from managers to external events. During the years of good performance, the letters to shareholders promote the actions of the management on the company's success. Letter to shareholder is an optional document that large corporations can choose to write for the shareholders and the public in general. They are typically supplied once a year with the annual report, however many companies simply choose to provide the relevant information in the annual report (labeled form 10-k in the United States) which is the primary research material of this study.

The annual reports and letters to shareholders have been a target for numerous studies into corporate culture and strategy, including the evolution of intellectual capital reporting (Vandemaele et al. 2005), corporate social responsibility reports of UK companies (Idowu and Towler 2004), and rise of firm actorhood in reporting (Bromley and Sharkey 2015). Bromley and Sharkey ran a content analysis of some 300 annual reports of firms between 1960s and 2000s to discover that, over time, firms increasingly use language that depict themselves as actors that act accordingly to their self-defined values such as social responsibility.

Annual reports are produced by the top management of the firm. As Livengood and Reger (2010) highlighted, the top managers beliefs were more aligned with the firm's actions than the beliefs of lower tier workers'. This supports the political-cultural approach of the firm as a battleground for various interest groups. I argue that, as the managers and shareholders of a firm constantly strive to control the firm as proposed by the principal-agent problem, the annual reports present a significant arena where the upper management can promote a vision of the market and the firm's place in it that is most suitable for them. This is supported by Poole's (2016) study where he conducted a keyword analysis on the letter to shareholders of two major american banks, Citigroup and Bank of America before, during and after the financial crisis of 2008. Poole discovered that during the times of crisis, the banks promote messages that assert vision and forward looking strategies of success while distancing the performance issues from managers to external events. During the years of good performance, the letters to shareholders promote the actions of the management on the company's success. Letter to shareholder is an optional document that large corporations can choose to write for

the shareholders and the public in general. They are typically supplied once a year with the annual report, however many companies simply choose to provide the relevant information in the annual report (labeled form 10-k in the United States) which is the primary research material of this study.

In chapter 2 I presented the general theoretical framework through which I approach the firm and its organizational identity, constantly being moulded by various interest groups. A major arena for this battle are the annual reports that are produced by the firm's top management. In the following section I will walk through the form of the annual report (called as form 10-k) in the United States.

3.1.2 - Form 10-K

To study how large organizations portray their industries, organizational identities, comparative differences and their change over time the researcher needs a corpus of documents that is both uniform, relevant and spread over multiple years. The annual report in 10-K form required by United States federal security laws to domestic companies should accommodate to these requirements. As I mentioned earlier, the central managing department for the forms is the U.S. Securities and Exchange Commission (sec.gov hereinafter "SEC"). The 10-K report is a filing that all US companies with assets more than 10 million dollars and 2000 owners must file to the SEC.

The 10-K form has four parts, first part must cover general business description, second part deals with the financial figures of the company, third covers the compensation for the company's executive branch and final fourth part miscellaneous items such as signatures. For the purpose of this study I will only deal with Part 1. Part 1 typically consists of some or all of following items:

General : A brief description of the company.

Industry Overview: A brief look into the market or markets where the company operates.

Products: One of the longer parts of Part I with detailed description of all the products the company has for offer.

Marketing and Sales: Discussion on how marketing and sales are have been handled for the past year.

Customers: Overview of the main customer segments of the firm.

Competition: Discussion of the major competitors of the firm and how it may affect present and future.

Research and Development: Look into the R&D operations of the firm, such as new product development.

Intellectual Property and Licensing: description of how intellectual properties and relevant laws may affect company business

Backlog: The current major orders placed for the company

Seasonality: discussion on how seasonality could affect sales (eg. christmas sales)

Employees: one or two lines about the number of employees

Environmental Regulations: brief description of how environmental regulation could

Risk Factors: typically a long description of various issues that could affect firm operations negatively. Some companies choose to disclose remarks about their competitive space in this part.

As my focus is on how the corporations discuss competition and how their organizational identity is potentially constructed through this discourse, I only focus on the “Competition” part of the report (underlined above). These are usually some two pages long descriptions. I should note that many of the parts in the annual reports presented above could be used for scientific study in social sciences. I will discuss these possible future research paths in the concluding part of this thesis.

3.2 - Method

3.2.1 - Big Data in Social Sciences

The modern world is awash with digital data (Hilbert and Lopez 2011). The amount of digital data humankind produces is also growing at an increasing pace and a great deal of this data is produced in text form, for example in official documents, news, ordinary interaction between people (such as messaging and ‘tweeting’ on Twitter) and digitization of previously analogous text sources using automatic processing technique such as optical character

recognition (OCR). This growing mass of data has attracted scientists to develop techniques to better make sense of it, from simple classification and clustering algorithms to more complex data wrangling methods such as artificial neural networks and also sophisticated data streaming approaches to big data like Hadoop and Mapreduce. The list goes on.

In social sciences, such methods have been gaining increasing attention as well (Burrows and Savage 2014), especially for handling large amounts of text. As Wiedemann (2013) says: *“For some years now computer-assisted text analysis is much more than just counting words. In particular, the combination of statistical and pattern-based approaches of text analysis, referred to as “text mining” may be applied to support established qualitative data analysis designs.”* Although computer-assisted techniques for text-mining have been widely used for decades in social scientific research, only recently has the computing power of computers developed to such level that more complex approaches are easily available to the wider scientific community. It should also be noted that this availability has sparked increasing interest to develop such methods. Wiedemann (2013) notes that qualitative social research has had a tendency to avoid mass data due to its self-conception as a counterpart to the quantitative research orientation. He further argues that, since the increasing availability of digital textual data and complexity of algorithms that can deal with such data has left a gap to the potential toolbox of qualitative social scientific research that should be filled.

Much of the advancements in handling big data is made possible by the increase of the computational capacity of humankind. At the micro level, this computational capacity refers to human constructed machinery’s capability to process information. The vast majority of this capacity is built on a natural phenomena whereby electricity, which is passing through specific kind of semiconducting material (most notably silicon), processes binary information. The basic unit of this information is called a “bit” which may assume one of two positions, 0 or 1. In other words, either an electrical impulse is passed on, or not. All modern computers are built following this principle.

In the modern capitalist society, most of this capacity is built using markets as an organizing device in production by firms that compete with each other in the “marketplace” to produce goods such as microchips. One purpose of this paper is to both demonstrate and develop a big

data method to study a particular aspect of the very process of competition that has built this capacity. This is achieved by using a massive digitized archive of annual reports of public companies. However this study merely scratches the surface of this immense pool of knowledge. As I enter somewhat unfamiliar waters in terms of methodology and data, I believe it is not ill-advised to tread lightly and use it on a relatively narrow target such as reports from only few companies.

3.2.2 - Topic Modeling

To help with the analysis of a collection of documents for this study, I apply a machine learning technique to classify the topical dimensions of the texts from the family of techniques called *topic modeling*. Topic modeling is a kind of unsupervised machine learning technique, a suite of algorithms, that is used for discovering abstract topics from a collection of documents (Blei et al. 2003; Blei 2012). Unsupervised learning means the algorithm does not require prior information of the data its modeling. It's a tool that provides a way to analyze large amounts of documents and find what they are about by clustering words together through an automated process. In topic modeling, a "topic" can be described as simply a cluster of words that share a statistical bond. In other words, they have a high probability of co-occurrence (Blei 2012). Topic modeling and other automated text manipulation tools have seen increasing popularity as the computer's computing powers increase and availability becomes ever more ubiquitous.

Mohr and Bogdanov (2013) argue that the most distinctive part of topic modeling is that: *"they provide an automated procedure for coding the content of a corpus of texts (including very large corpora) into a set of substantively meaningful coding categories called "topics." The algorithms can do this with a minimum of human intervention, and this makes the method more inductive than traditional approaches to text analysis in the social and human sciences."* This feature makes topic modeling a specifically potent exploratory technique for text analysis.

The topics in topic modeling, that is the clusters of words inferred from a selection of documents with topic modeling algorithms, can be described as "frames" (Fligstein et al.

2014; Goffman 1974) that help explain events through actors' prior cognitive capacities. In the case of annual reports, these frames are the views of the upper management about the company, its financial situation, competition and so on, that they choose to comment upon. Goffman uses the term “primary framework” to refer to the way intelligent lifeforms render meaningless aspects of a scene into something meaningful (Goffman 1974, p. 21).

Dimaggio et al (2013) argue that topic modeling satisfies four critical conditions of text analysis: Explicitness (1): the data and methods are available for others to reproduce the study. Automation (2): the method accommodates to even large volumes of data. Inductiveness (3): the method permits the researcher to discover the structure of corpus before imposing their priors on the analysis. And finally, Capability to recognize the relationality of meaning (4): the topics are formed through their co-occurrence in the corpus. This means that the same word can be part of different topics and semantic structures.

3.2.3 - Latent Dirichlet Allocation

In this study I will be using a topic modeling algorithm called latent dirichlet allocation (hereinafter "LDA") which is, due to it's relative simplicity (Blei 2012), one of the most commonly used topic modeling algorithm. The mathematical representation of LDA can be found from, for example, Blei 2012. For this paper, I will present the model in more common sense terms. LDA is a member of a field called probabilistic modeling. Probabilistic modeling assumes the data to contain some hidden variables which can be discovered from its observable, visible variables. In the case of LDA, the visible variables are the words of the documents, and the hidden variables are the topics in those documents. For conducting LDA, one must first process the documents into a vectorized mathematical form known as bag of words, in order for the computer algorithm to be able to handle it. Bag of words form means that the program does not know the order of words in individual documents, only their amount per document. The bag of words form has some important implications for the study of text (Fligstein et al. 2014):

- 1: the topics may be dispersed throughout the text and a single sentence may have multiple topics encoded to it.
- 2: the context of the topics may be left unacknowledged, for example a document could be strongly opposed to a topic but the model may only recognize that the discussion is about the topic, not whether it's positive or negative.

The relational nature of words is a key part of the structuralist approach to language which assumes that the meaning of words arise, not from the words or symbols themselves, but from their relation to other words or symbols. Dimaggio et al. (2013) propose that this helps with semantic validation of different topics; When same word is assigned to different topics, different meanings should be evident.

The way data is chosen and prepared for the algorithm involves a large amount of decisions by the researcher that will affect the outcome. Some of these choices may be arbitrary. These choices involve the pre-parameters of the algorithm and the choice of input documents from which the model is created and to which the model is used. These choices affect the way I envision “the social” in this study. As Mützel (2015) notes, while LDA, like all topic modeling techniques, is described as “automatic process”, it's far from it since the amount of work before a model is created, is immense. As a crucial note to this study, Mützel calls for more transparency in how we arrive at the data, be it qualitative or quantitative.

LDA has seen frequent employment in social sciences in the recent years: Dimaggio et al (2013) applied it to analyzing newspaper coverage on a specific topic over a timespan of ten years with a document corpus consisting around 8000 articles. Fligstein et al (2014) used LDA to investigate why Federal Reserve failed to predict the financial crisis of 2008 by analyzing transcripts of the meetings of its central decision-making body, the Federal Open Market Committee over 10 years. Koltsova and Koltcov (2013) analyzed posts from a Russian blog platform LiveJournal using LDA and discovered, for instance, that the users distribute their attention fairly evenly among private or recreational topics and social or political topics. Miller (2013) used LDA on the official records of Qing dynasty China to model the records' topical dimensions and study their relationship with the varying periods of social unrest in the region.

LDA, while being widely popular, also has some issues that need to be addressed, if used for scientific research purposes. For example, Koltcov, Koltsova, and Nikolenko (2014) found out that the topics emerging from multiple runs on same dataset varied considerably. As the results also differ to some degree depending on the parameters given to the model, results from topic modeling are not suitable to be used as evidence for scientific research. They should instead be understood as an exploratory device. This is supported by Chang et al. (2009) who conclude, after conducting human experiment on LDA's topic validity, that LDA is indeed suitable for exploratory purposes, so long as its offside are kept in mind. DiMaggio et al. (2013) propose that LDA and other topic modeling techniques should be thought of as a "lens" which helps to analyze large amounts of textual data; It is up to the researcher to visually choose the appropriate lense (that is, the parameters for the algorithm such as the number of topics) with which an approximately optimal clarity is achieved (for more on topic stability in topic modeling see Greene et al., 2014).

Jacobi et al. (2016) note that the topics in LDA, inferred from the co-occurrence of words, means that the not only topics as theoretical concepts but they also might represent other patterns of co-occurrence such as particular events or talking styles. She notes that, although other scholars such as Dimaggio et al. (2013) treat these topics simply as frames, one should approach the question of what they represent simply as an empirical question. As I analyze the outputs in the upcoming sections, I will heed this advice and approach the emerging topics as a product of word co-occurrence. The topics that are inferred from the text using LDA can be also described "features" of the modeled text to follow the common nomenclature of computer science. These features are the co-occurring words that the algorithm places under a topic.

To summarize, LDA should be considered as a more inductive and sensitive approach to textual analysis when compared with traditional methods such as a simple word frequency analysis. As Dimaggio et al. (2013) summarize: "*With topic models, researchers can discover new patterns in their text data and analyze much larger collections than is possible by hand.*" In this thesis my focus will be on using topic modeling to reveal temporal and relational changes in how competition is discussed in annual reports.

3.3 - Use of the Data and Method in this study

As already mentioned, LDA takes all the words in a document and treats them as a "bag of words", in other words, the order of words is irrelevant in the document. A document can be of any length, although short documents, such as Tweets from the microblogging platform Twitter, are found to be somewhat harder to model than longer texts because of the document level word co-occurrence levels tend to be too low (Yan et al. 2013). A document in LDA is an abstract entity, meaning that the algorithm can take complete documents, individual paragraphs or even sentences from the documents to model from. Ideally, one would use complete documents as input when studying their temporal topical variations, however in this study the number of documents is somewhat low, so I opt to input paragraphs instead to get a more granular view on these documents.

After ingesting the document corpus as bag of words, the algorithm then follows a multi-stage process (Blei 2012): first randomly choose a distribution over topics. Then for each word in the document randomly choose a topic from the distribution over the topics defined previously. Then randomly choose a word from the corresponding distribution over the vocabulary. This process is then iterated a number of times as defined in the parameters given to the algorithm. Finally, all the words are assigned to one or more topics inferred from the documents. Ultimately, the topics and their interpretability are derived from the statistical structure of language (Blei 2012). A "document" here is an abstract entity meaning a piece of text, which can typically be either the whole document, such as an annual report, a paragraph from such document or even a sentence. As mentioned previously, shorter documents, such as sentences, are difficult to model from but also longer documents may be problematic for finding specific topics from the texts.

As it is with most if not all machine learning algorithms, the process by which the LDA algorithm produces a matrix of topic distributions is a kind of "black box" in the sense that the researcher can't know the exact way the output was created and due to the probabilistic approach (the algorithm starts with a random choice of topics from which it infers future

probability distributions). There is also an inherent randomness with the results. This means that the researcher has to ground their findings with a triangulation strategy, most typically achieved with a close reading of parts of the original documents and good understanding of the subject of study. In the case of this research study, I will accompany LDA with samples to help understand how the topic is discussed locally in the documents.

For this paper I will be using the Python programming language for processing and analyzing the documents. Python has become very popular language for data analysis and machine learning due to its high readability when comparing to other languages. Before running the text corpus into the model, several pre-processing steps are commonly taken to increase the quality of the output so as to avoid the "*garbage in, garbage out*" problem commonly observed when running machine learning algorithms on data. I will describe the process in layman's terms here and the full code, as well as links to data, are presented in the Appendix in order to assure easier scientific replicability of the results.

First, the text corpus was loaded and all individual characters made lower case. Then some of the most common words in English (typically called "stopwords") were removed from the documents (the list of stopwords is derived from the NLTK natural language toolkit package). After this a widely used stemming algorithm (Porter, M. F. 1980) was used to find the stem of the words so that the different variations of the same word will not be regarded as different words (for example words "ability" and "abilities", would both be stemmed to "abil").

I created the model using the whole "Part I" from every annual report from 2004 to 2016 produced in 10-k form for both Intel and AMD separately. The total number of these documents were 26 (13 for both firms). For the model, I had to choose the size of documents from which the LDA algorithm would calculate topic distributions and the best topics seemed to be produced by splitting the documents into paragraphs, since the whole document as input would have made it too difficult for the algorithm to find good word co-occurrences from the relatively small amount that were used in this study. This model was then used to analyze the "competition" subsection from the annual reports on two corporations: AMD and Intel on an yearly basis. Since the LDA transforms the documents into a mathematical representation, it

is possible to inspect the relative quantities of different topics that are visible in a chosen document. As I have documents that cover a period of time, I am then able to look how different topics vary in relative presence over this time. Similar approaches have been used in many of the social science studies that made use of this method, which I have discussed earlier (for example, DiMaggio et al. 2013, Fligstein et al. 2014).

Following the widely adopted approach (for example Fligstein et al. 2014) I ran the model with different amounts of topics and then qualitatively assessed the optimal topic number for my corpus. I ended up with 9 topics for AMD and 12 for Intel. Since I am using this model to focus on just one part from the annual reports (Competition -subsection), this level of sophistication seems appropriate. After the model is created and used on the competition subsection for both firms, I visualize the topical dimensions and relative distribution on a year by year basis. This will allow me to see how different topics vary over time in these documents.

LDA is an exploratory technique used to impose order on large text data. As such it should, theoretically, be an ideal research method for analyzing large masses of textual data in social sciences since, as defined by Ragin and Amoroso (2010), the very purpose of social scientific research is to “*find order and regularity from the complexities of social life*”. In this thesis, the use of LDA has two purposes. Primarily, to simplify the large corpus of data used for this study, and reveal its underlying topics and themes that could otherwise be left unexposed. LDA is a method that transforms text into a mathematical form, and therefore allows the researcher to visualize some features of text in ways that are impossible in more traditional methods. The other purpose for use of LDA in this thesis is to explore new uses for this relatively new method and to test it in a novel way.

4 - Results

4.1 - Foreword to results

As I have proposed in the previous chapter, LDA can be used as an inductive research tool which helps the researcher to find interesting features from the text that would be much more difficult to discover with more traditional text analysis methods. In this chapter I show the topics that are of most interest to this study and discuss how they should be understood and how they change over the years. In Chapter 5 - Discussion, I will make general remarks how these topics and their change can be used to understand competition as a facilitator of organizational identity. The algorithm automatically numbers the topics, however I will also give the topics names to help make it easier and clearer to refer to them.

4.2 - AMD

For AMD, the best results were achieved with 9 topics. This resulted in a set of topics which can be seen in Appendix. In this part, I look at four topics that are of specific interest to this study and finally how they vary over the observation period (figure 1).

4.2.1 - Intel as aggressive competitor -topic

This topic, seen on the right (table 1) is about how AMD sees Intel as a company that uses aggressive methods and its superior resources and position to control competition. One should keep in mind this is not so much about Intel, but about how AMD portrays Intel in their reports. The topic is also highly discussed throughout the observation period (see figure 1 below). AMD clearly portrays Intel as the dominant player in the semiconductor industry which, as discussed in introduction chapter, is how Intel has generally been seen by analysts, government legal bodies and other such interest groups. Below I have picked a quote from AMD’s 2015 report to illustrate how this topic is discussed in the reports with words from the topic underlined:

“Intel Corporation has dominated the market for microprocessors for many years. Intel’s market power and significant financial resources enable it to market its products aggressively, to target our customers and our channel partners with special incentives and to discipline customers who do business with us. “ (AMD 2015.)

This topic shows how AMD uses Intel as a way to construct its own identity. As I argued earlier, when a company is described with an adjective such as “high quality” or indeed “aggressive” as shown here, it can only be understood when compared with some other company that is not “high quality” or “aggressive”.

As AMD has a long history of accusing Intel of unfair practices in the semiconductor business and the two firms have been solving disagreements in the courtroom from time to time, it should be interesting to see how this topic changes over time. Some of these disputes

2
market
intel
product
microprocessor
aggress
price
research
greater
develop
financi
resourc
lower
competit
technolog
manufactur
domin
custom
substanti
advers
compani

table 1

were settled in 2009 with a joint statement from both companies stating all disagreements have been solved: "*While the relationship between the two companies has been difficult in the past, this agreement ends the legal disputes and enables the companies to focus all of our efforts on product innovation and development.*" (Intel.com 2009). Interestingly AMD does not stop portraying Intel as an aggressive player in the industry as seen from figure 1. The figure gives an image of how this topic is discussed in the reports throughout the observation period:

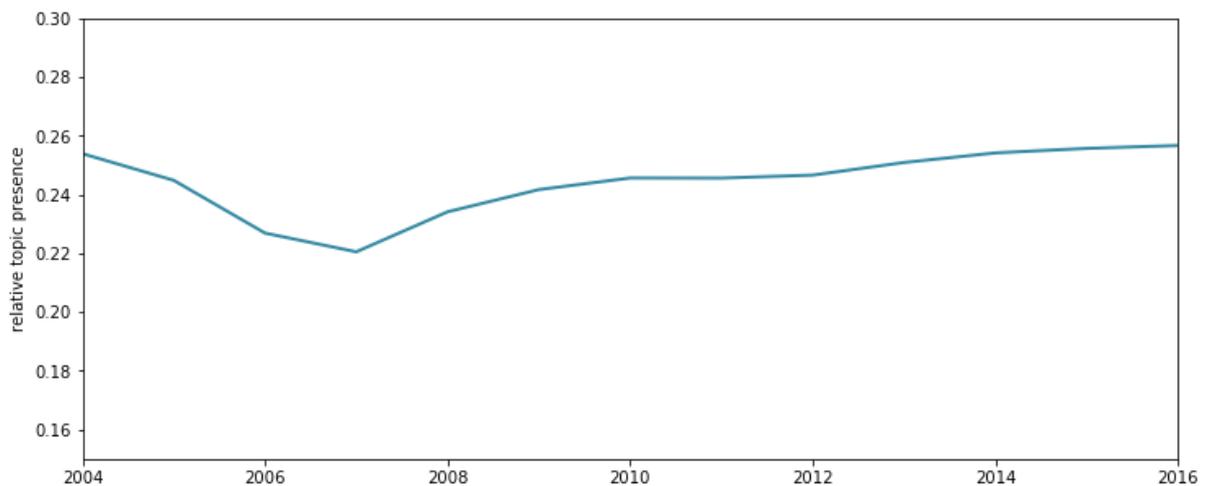


figure 1, Intel as aggressive competitor -topic

As the figure shows this topic consists of about fourth (22% - 26%) of the whole sub section's topical space throughout the years. It means AMD invests a considerable time to discuss Intel and portray it as an aggressive player in the semiconductor industry. Through this we can draw a conclusion that Intel is a very important part of how AMD constructs its perception of the semiconductor business and ultimately itself. As AMD portrays Intel in this way, it creates an image of itself as a smaller and more fair player who has to comply with Intel's unfair practices. It is worthwhile to note that this topic (as seen on figure 1) barely loses presence even though, as was mentioned in the introduction, AMD officially stated its intentions to move out from the old AMD/Intel mindset in its strategic operations in 2011. It would seem such drastic changes to organizational identity are not easily undertaken.

4.2.2 - Intel and component market -topic

The algorithm detected another, somewhat more mellow topic regarding Intel. This topic discusses Intel as the dominating rival for AMD. The topic includes words such as “microprocessor”, “motherboard”, “graphic” and “chipset” which indicate that the company is a significant competitor in many different product segments that deal with computer components. For further reference I name this topic “Intel and component market” -topic. Below is a quote from an annual report that features many of the words assigned to this topic:

“As a result of Intel’s dominant position in the microprocessor market, Intel has been able to control x86 microprocessor and computer system standards and benchmarks and to dictate the type of products the microprocessor market requires of us. Intel also dominates the computer system platform, which includes core logic chipsets, graphics chips, motherboards and other components necessary to assemble a computer system.” (AMD 2011).

In this topic, AMD portrays Intel as a player which can determine the rules of the field, in the cultural-historical parlance, the conceptions of control that determine what is at stake in the field of semiconductor business.

The previous topic, “Intel as aggressive competitor”, describes Intel in much more negative tone, while this one is more neutral. The interesting part of this topic is how AMD argues that Intel is able to dictate the rules of the semiconductor field. In the sociological parlance that I introduced earlier, we can argue that AMD sees Intel in a position on the market where it can make its conception of control the dominating logic in the field. Defining standards and benchmarks and dictating the type of products that the market requires of producers is exactly

5
intel
microprocessor
system
comput
standard
market
domin
manufactur
motherboard
chipset
includ
control
industri
compani
graphic
platform
posit
compon
facto
logic

table 2

what Fligstein (2001) argue the dominating conception of control to affect. Below the topic's annual variation is displayed:

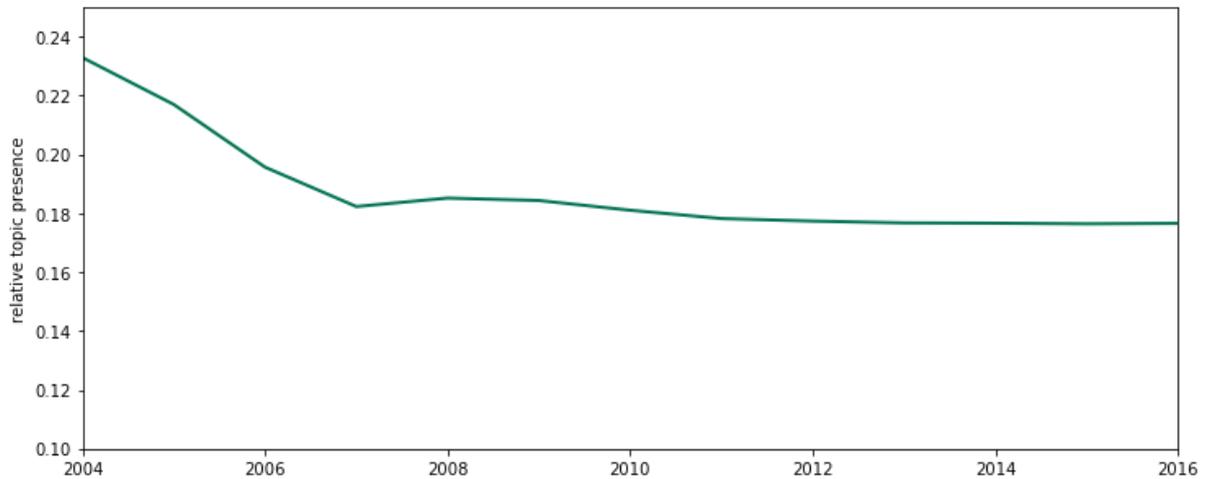


figure 2. Intel and component market -topic

The figure shows this topic to slowly lose presence between 2004 -2007, then maintaining a stable presence. As I discussed earlier, when a firm defines itself in terms of differences and similarities to other companies, they are addressing two important aspects of survival in modern market economy: competitiveness and legitimacy. AMD clearly draws legitimacy from its higher status competitor Intel by focusing on it extensively and constantly, as seen from figure 2, in their annual reports. As organizational identity was earlier in this study defined as that which is most central, enduring and distinctive in the organization, I argue that this topic well adheres to these three qualities: it endures throughout the years, it is central in its high relative presence, and it is how AMD distinguishes itself from a competitor.

4.2.3 - Graphic chipset market -topic

The topic which the algorithm issued number 7 is about graphic chipset market which are also the top three words to this topic. Graphic chipsets, which are most commonly known as GPUs or the graphical processing units, are computer components that process the visual inputs and outputs of a computer. This topic is no doubt important for AMD as it acquired ATI Technologies incorporated in 2006 in a move intended to place AMD firmly into the GPU market. Below is a quote in which AMD talks about this topic (note that the word “GPU” itself was not included in this topic which was because I set the minimum character length to four in the text preprocessing stage because I assumed short words to be irrelevant to the topics. On a hindsight, this may have not been an optimal choice as three character long words such as GPU does provide a lot of semantic substance):

“In the chipset market, our competitors include suppliers of integrated graphics chipsets. PC manufacturers are increasingly choosing to use integrated chipsets, particularly for notebook computers, over discrete GPUs because integrated chipsets can cost significantly less than discrete GPUs while offering acceptable graphics performance for most mainstream PC users. Intel Corporation manufactures and sells integrated graphics chipsets bundled with their microprocessors and is a dominant competitor in this market.” (AMD 2010).

7
graphic
chipset
market
integr
competitor
discret
intel
manufactur
could
microprocessor
domin
includ
sell
offer
perform
competit
nvidia
supplier
mainstream
user

table 3

Even in the business for GPUs the prevalence of Intel as the main competitor is prevalent. Also notable is the presence of another prominent GPU manufacturer, Nvidia in the topic, however it resides much lower in the list indicating much lower topical importance of the firm for AMD. In the next figure I investigate how this topic changes in presence over time.

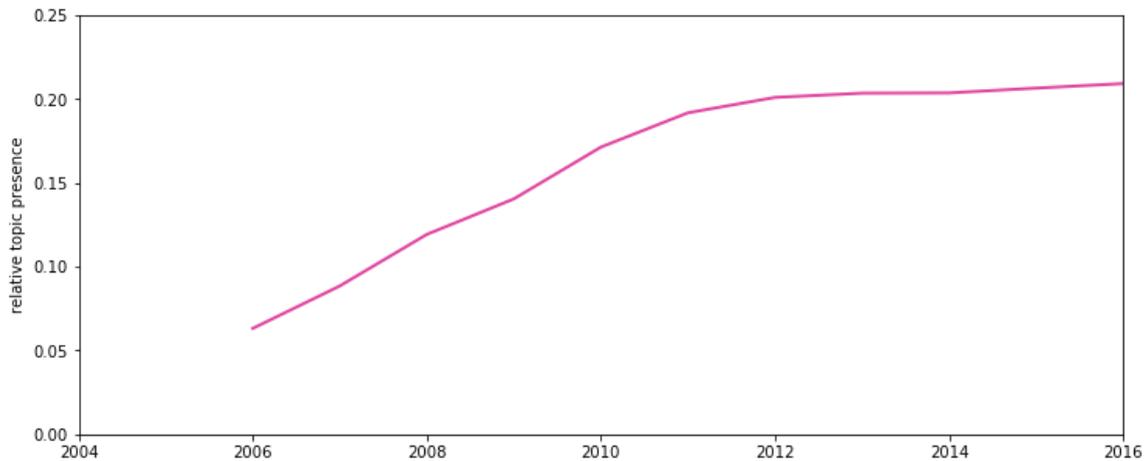


figure 3. Graphic Chipset -topic

As the figure 1 shows, Graphic Chipset -topic emerges in 2006, the very same year AMD acquired ATI technologies. The topic rises in frequency until 2012 after which it occupies a constant presence in the reports. This indicates AMD shifting much of its focus on GPUs. Interestingly the Graphic Chipset -topic rises in presence gradually between 2006 to 2012 and then retains a stable and high presence. This figure shows the creative process by which organizational identity transforms through time. Such changes seem to take time for an organization as seen from the figure 3. In 2004 and 2005 AMD hardly identified as being a GPU manufacturer or having much stake in the market at all. From 2011 onwards the topic however holds some 20% presence in the report, indicating a drastic transformation of AMDs identity.

4.3 - Intel

For Intel I chose two topics, out of the twelve, that seemed to be of most interest to our study. Just like for AMD, the twelve topics can be seen in Appendix while here I only focus on these two.

4.3.1 - Product segments and market -topic

The first topic from the algorithm seen on the right is a very general topic which demonstrates Intel’s awareness to product segments, competition and markets where it operates in.

“Many companies compete with us in the various computing, networking, and communications market segments, and are engaged in the same basic business activities, including R&D. Worldwide, these competitors range in size from large established multinational companies with multiple product lines to smaller companies and new entrants to the marketplace that compete in specialized market segments. “ (Intel 2007.)

As the quote above demonstrates, this topic is very general in tone. Intel talks about its competitive environment in a way that does not specify competitors. As Intel is an industry leader in semiconductor business, it is interesting to see how general and unspecific the talk about competition is when compared with AMD. This holds true for all of the topics that were produced by the algorithm (see Appendix) as no firm has a significant position in the topics.

1
product
segment
competitor
market
compani
competit
busi
compet
offer
comput
commun
softwar
devic
larg
marketplac
increas
emerg
numer
multipl
establish
industri

table 4

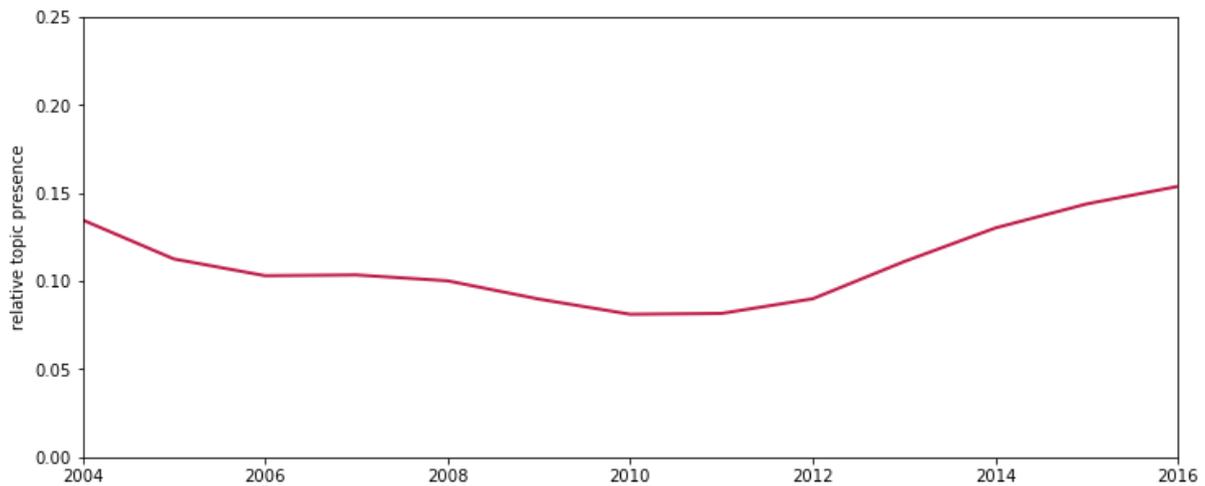


figure 4, Product segments and market -topic

Figure 4 displays how this topic changes over time. There is a slight build-up from 2012 onwards in this topic, which indicates Intel’s increasing focus on new technologies and market segments it could pursue. Below is a more recent quote which demonstrates this discourse:

“As technologies evolve and new market segments emerge, the boundaries between the market segments that we compete in are also subject to change. Intel faces significant competition in the development and market acceptance of our products in this environment.” (Intel 2016.)

One way to understand this topic is to look at how Intel has performed in its core product segments, namely microprocessors. As many analysts note, Intel has been dominating in the core market of consumer grade microprocessors for years now as AMD has not been able to produce the same price to performance ratio as Intel (Ubilla 2017). This position has given Intel space to focus more on possible new markets and product segments.

4.3.2 - Industry standards -topic

This topic is about competing standards from different firms that affect Intel’s competitive position. This topic is particularly interesting when compared with AMD’s Intel and component market -topic which revolves around AMD accusing Intel of power to set standards in the industry. Below is a quote that tells a different story from Intel’s perspective:

“Our products compete with products developed for similar or rival architectures and with products based on the same or rival standards. We cannot predict which competing standards will become the prevailing standards in the market segments in which we compete.” (Intel 2006.)

Intel clearly does not want to admit it has power to define rules of the market in the way AMD accuses it to have. Another quote below further illustrates how Intel understands the standard setting process:

“We frequently participate in industry initiatives designed to discuss and agree upon technical specifications and other aspects of technologies that could be adopted as standards by standards-setting organizations. Our competitors may also participate in the same initiatives and specification development. Our participation does not ensure that any standards or specifications adopted by these organizations will be consistent with our product planning.” (Intel 2010.)

There are indeed some striking differences between how AMD portrays Intel as an aggressive player who manipulates standards and rules of the market as it sees fit, and how Intel itself sees itself as simply a participant in the wider process where standards and rules are set in co-operation between a plurality of actors in the market.

11
standard
specif
industri
particip
product
adopt
organ
initi
market
intel
softwar
develop
advantag
system
platform
comput
usag
offer
oper
compet
abil

table 5

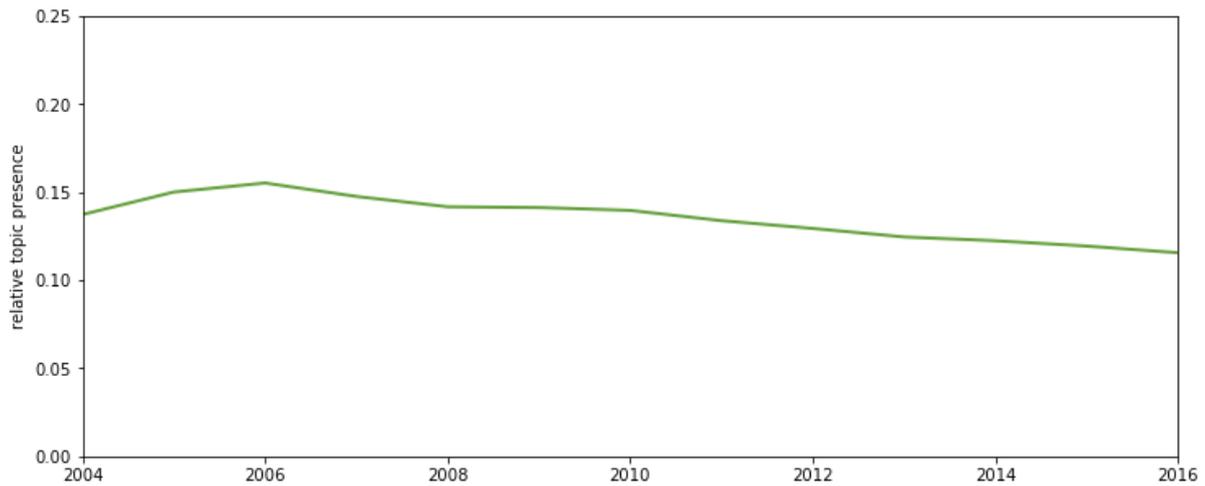


figure 5, Industry standards -topic

As can be seen from the figure 5, this topic hardly changes throughout the years, going from some 15% relative presence down to 12%. Due to the probabilistic nature of the LDA and the somewhat small amount of data fed into it in this study, such small changes are not significant. In general, Intel's reporting about its competitive environment has been relatively stable through the observation period of 2004-2016 (see Appendix). This may be due to its position as an industry leader in a fairly stable market. New entrances and changes hardly happen, even in a decades time. A further research with more companies could give some insight into whether dominant position makes a firm perceive its competitive environment more stable or not. I would assume that dominant firms don't need to change their vision of the market and itself unless a challenger rises to substantially threat its position. Such claim would warrant further research with wider selection of companies in different markets.

5 - Discussion and conclusion

5.1 - Review of findings

The most distinctive feature that arose from the reports was the degree at which AMD focused on Intel in its report and how Intel's discussion was much more general in its tone. I argue that this is mainly due to their different positions on the market, in the cultural-political framework we can call AMD the challenger firm and Intel the incumbent. Due to their different positions in the market two companies clearly perceive it in significantly different ways. This position then affects the firm's organizational identity as challenger firm seeks to position itself next to the incumbent. Intel as an incumbent firm seems to have a tendency of portraying itself as a firm that constantly needs to adapt to changing competitive environment to which it has no power over. This is in stark contrast to AMD's portrayal of Intel as a "hegemon" of semiconductor business, changing standards and industry practices whenever it suits them.

In this paper I sought to explore how organizational identity is constructed discursively through firm's portrayal of their competitive environment. Albert and Whetten (1985) defined organizational identity as that which is most central, distinctive and enduring to the organization. I then argued that the annual reports could be used to research these dimensions through the firm's portrayal of its competitive environment. When a company describes its competitors, it defines itself in the process as something different. This was particularly evident in AMD's reporting whereby it constantly described itself as the smaller challenger to Intel. Another notable finding was the relative stability of these topics, except for the Graphic chipset -topic from AMD. This supports the earlier findings (Tripsas 2009) about the difficulty of changing organizational identity, even in the face of changing technological landscape which should be particularly strong for firms operating in the technology sector.

The use of LDA may not have been the optimal choice for this particular study as it lacked the capability to trace small discursive changes in language in the annual reports. The topics themselves were not as easily identifiable as I would have hoped to help understand the way

these companies discuss competition and their differences. As the amount of documents that I used in the study was relatively low, the algorithm may not have been able to form topics that were more easily readable. This stands as an important reminder that machine learning algorithms should be used with caution and preferably with large amounts of data as the more there is to learn from, the better the resulting models for machine learning tend to be. I would argue that there is value to using these emerging methods first with smaller amounts of data since it makes it easier to understand how a particular topic was created. Even with such small amount of input text, the algorithm was able to simplify the text in such way that some conclusions about organizational identity and competition could be made.

5.2 - Discussion on methodology

The main idea of this study was to analyze how organizational identity is constructed discursively in annual reports of companies with a focus on the way they portray their competitive environment. To help observe the temporal changes in the reports, I created a model of the text data with technique called LDA. It creates an abstraction of the text by grouping similar words together which are then understood as different topics that are discussed in the documents. LDA, and other topic modeling techniques, act as an interesting bridge between the qualitative and the quantitative research tradition as it creates a mathematical abstraction of text that can be visualized in a mathematical form. The old division between qualitative and quantitative research has been under some scrutiny as all research tends to have both aspects in them (Töttö 2004). Topic modeling as a research tool illustrates this argument since it can hardly be classified as purely one or another.

The largest issue that I discovered using the method in this study was the loss of sensitive touch with the text that is usually achieved with more traditional qualitative approaches such as discourse analysis. In these traditional methods that are widely used in social sciences, the researcher approaches text from a very intimate distance, drawing insightful remarks from even the smallest changes in semantic dimensions of individual sentences or words. LDA creates a model of the text that is powerful when analysing large amounts of text but simply can't, on its own, reach such delicate features of textual data. In last chapter, I provided parts of text to aid in understanding how a topic is typically discussed in a report. There may have

been small and interesting changes in the way these topics are discussed in the reports over the years which, due to the generalizing nature of LDA, is lost in the approach used in this study.

LDA as a method is geared towards processing large, or even humongous amounts of textual data, which hardly was the case in the study presented here. After all the 26 documents from two companies is a relatively small collection of data to be used for a technique geared for processing huge amounts of it. Such a limited amount of data could have been analyzed with methods that did not need to use of such black-box method that, while giving user the benefit of machine learning techniques as a clustering tool of thematic features of the text, also derives the researcher from clearly understanding how the results were obtained. The main issue of using LDA in this study was the method's inability to detect small changes in discourse over the years. As words are clustered through co-occurrence, the resulting topic do not give us any insight into small changes in how a topic was discussed between years.

I have used a relatively new approach to analysing textual documents to explore how firm's top management discuss competition in a high-tech sector in their annual reports. Annual reports are a potential treasure trove of interesting data for economic sociologists about the basic unit of capitalist production: the corporation. In this study, I have only scratched the surface of what the combination of big data methods and a massive pool of these official documents are capable of. An example of possible future directions for studying annual reports would be to use document similarity metrics (for example calculating a Hellinger distance between annual reports of different companies, which measures distances between different probability distributions) in combination with LDA and the theories on organizational identity to measure differences between multiple firms that operate on same market or markets and map entire industries with such metrics.

Machine learning, and artificial intelligence in general, are the great buzzwords of this decade. Time will show how they will change human societies and how they affect scientific research as well. It would be hard to not see the possibilities of these not used widely in the years to come for deepening our understanding of reality. Social sciences have traditionally been somewhat conservative in the use of such cutting edge methods. As my study has

shown, modeling social interaction in a way that takes into account the delicate semantic dimensions of human reality is no simple task and it requires considerable effort on behalf of the user to understand how the machine came to the resulting model. Only way to deepening our understanding such methods is by using them with a critical mindset, duly noting the shortcomings of the method as well as its strengths.

Some interesting future paths to explore with LDA would also be to combine it with theories that are more refined for extracting and analysing certain dimensions or aspects from text. For example, the theory of justification, originally presented by Laurent & Thevenot (1996), could potentially be combined with LDA to discover how different “worlds” interact in large textual databases. The theory on justification, developed by Laurent Thevenot and Luc Boltanski (Boltanski & Thevenot 2006) offers an interesting way to analyze the moral dimensions of arguments used in public argumentation (Luhtakallio & Ylä-Anttila 2011). Boltanski and Thevenot propose that all discourse aimed towards a “public” are fundamentally based on a moral “world” of understanding. In situations where a dispute has arisen and use of violence is not an option, people must rely on argumentation to solve their differences. The scholars propose that these arguments must be based on commonly shared moral conventions that have shaped through humanity’s history of reason. The authors originally proposed six different worlds: the worlds of Market, Industry, Fame, Inspiration, Citizenship and Home. These worlds are not final or universal, a point that was emphasized when Boltanski and Thevenot added a seventh world, the world of Ecology to their framework. According to the authors, these worlds are bound on collective cognition of humans, the shared social world that is under constant reflexive transformation. Topic modeling could benefit drastically from the use of such powerful conceptual tool for framing the topics.

LDA produces very complex models and I only presented one, fairly common, use of the model by discussing how the topics vary over time. There are ways to visualize LDA in much more complex detail using more refined computer technologies than simple text, for example HTML code. I believe the way scientific research is presented to the scientific community and wider public will change drastically in the future by making use of technologies that allow very complex models to be more easily investigated. In the case of LDA, for example,

there is a visualization package called LDAvis that visualizes the topics, their semantic distances between each other on a two dimensional space and the frequency of the terms on a topic as well as its exclusivity to the topic (Sievert & Shirley 2014). One of the motivations behind LDAvis is to help interpret the topics produced by LDA as the topics have been noted to be often difficult to interpret (Chuang et al. 2006).

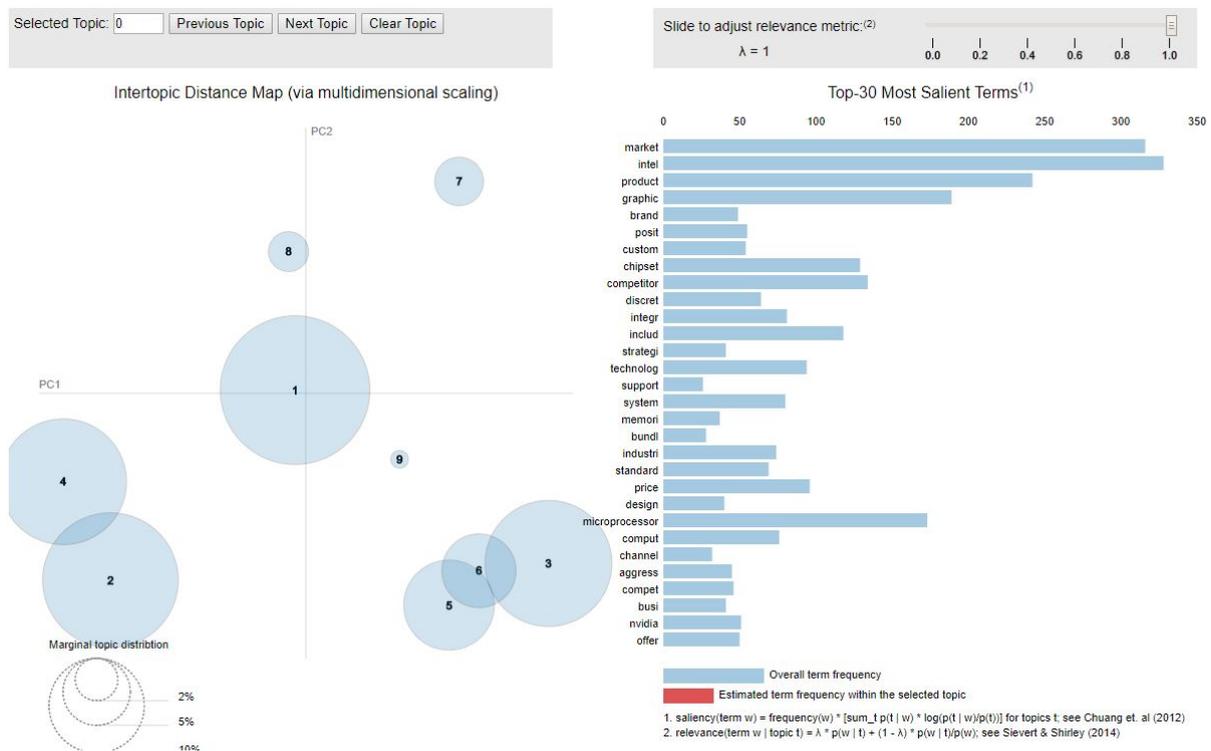


figure 6, a visualization of LDA using HTML

Figure 6 is a screenshot of the HTML output from LDAvis visualization package using the model that used in this study for AMD consisting of the 9 topics created by the algorithm (shown in the Appendix). A possibility to adjust relevance of words to specific topics gives considerable benefits to interpreting the topics.

5.3 - Conclusion

Markets are the great organizing principle of the modern economy as they place firms into competition between one another to produce commodities, ideally at ever higher quality and lower price. As modern capitalist corporations themselves are a battleground for various

interest groups such as workers, managers and owners, these groups use language to make sense of what their organization is, and what their place is in the wider society. Every company has their own organizational culture, that is conventions and conceptions of what the firm is all about. This is called the firm's organizational identity. In this study I have sought to explore how two firms' top management use competition as a conceptual tool to construct their own organizational identity as something that distinguishes themselves from their competitors. I also argued that only a distinctly sociological theory of action can make sense of what happens when organizations face pressure to change their organizational identity. As the two firms in this study discuss competition, the incumbent firm, Intel seemed to have a tendency to portray themselves as passive agent that had to co-operate with other firms in the market to create rules and agreements that make fair competition possible, while AMD constantly used Intel as a the facilitator of its own identity as the fair player that protects consumer choice against the monopolistic practices of Intel. In the future, expanding this study into a plurality of markets and firms could produce a much more refined picture of how competition affects organizational identities of both incumbent and challenger firms and what sort of commonalities emerge from these sociologically fascinating dynamics.

6 - References and Appendix

6.1 References

Ahrne, G., Aspers, P., & Brunsson, N. (2015). *The organization of markets*. *Organization Studies*, 36(1), 7-27.

Albert, S., & Whetten, D. A. (1985). *Organizational identity*. *Research in Organizational Behavior*, 7, 263-295.

AMD.com, AMD. *Worldwide antitrust enforcement*. Retrieved from <http://www.amd.com:80/en-us/who-we-are/corporate-information/competition/antitrust-enforcement>

Intel.com (2009). *AMD and intel announce settlement of all antitrust and IP disputes*. (2009). Retrieved from https://www.intel.com/pressroom/archive/releases/2009/20091112corp_a.htm

Betanews (2007). *AMD CEO: Intel is a monopoly, microsoft isn't*. Retrieved from <https://betanews.com/2007/06/21/amd-ceo-intel-is-a-monopoly-microsoft-isn-t/>

Anand, V., Joshi, M., & O'Leary-Kelly, A. M. (2012). *An organizational identity approach to strategic groups*. *Organization Science*, 24(2), 571-590. doi:10.1287/orsc.1120.0752

Ashforth, B. E., & Mael, F. A. (1996). *Organizational identity and strategy as a context for the individual*. *Advances in Strategic Management*, 13, 19-64.

Berger, P. L., & Luckmann, T. (1991). *The social construction of reality: A treatise in the sociology of knowledge* Penguin UK.

- Blaudziunas, R. (2016). Information about article authors. *Engineering Economics*, 27(2), 420-437. doi:10.5755/j01.ee.27.2.14867
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84. doi:10.1145/2133806.2133826
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 1022-1993.
- Blei, D., Carin, L., & Dunson, D. (2010). Probabilistic topic models. *IEEE Signal Processing Magazine*, 27(6), 55-65. doi:10.1109/MSP.2010.938079
- Boivin, C., & Roch, J. (2006). Dominant organizational logic as an impediment to collaboration. *Management Decision*, 44(3), 409-422. doi:10.1108/00251740610656287
- Boltanski, L., & Thvenot, L. (2006). *On justification: Economies of worth* Princeton University Press.
- Bromley, P., & Sharkey, A. J. (2015). "Casting call: The expansion of actorhood in U.S. firms, 1960-2010". *Academy of Management Proceedings*, 2015(1), 15743. doi:10.5465/AMBPP.2015.15743abstract
- Burrows, R., & Savage, M. (2014). After the crisis? big data and the methodological challenges of empirical sociology. *Big Data & Society*, 1(1), 2053951714540280.
- Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. Reading tea leaves: How humans interpret topic models. ().
- Chuang, J., Gupta, S., Manning, C., & Heer, J. (2013). Topic model diagnostics: Assessing domain relevance via topical alignment. Paper presented at the Proceedings of the 30th International Conference on Machine Learning (ICML-13), 612-620.

- Coase, R. H. (1937). *The nature of the firm*. *Economica*, 4(16), 386-405.
- DiMaggio, P., Nag, M., & Blei, D. (2013). *Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding*. *Poetics*, 41(6), 570–606. doi:10.1016/j.poetic.2013.08.004
- Evans, S. K. (2015). *Defining distinctiveness*. *International Journal of Business Communication*, 52(1), 42-67. doi:10.1177/2329488414560280
- Fligstein, N. (2001). *The architecture of markets: An economic sociology of twenty-first century capitalist societies*. United States: Princeton University Press.
- Fligstein, N., & Dauter, L. (2007). *The sociology of markets*. *Annual Review of Sociology*, 33(1), 105-128. doi:10.1146/annurev.soc.33.040406.131736
- Fligstein, N., & McAdam, D. (2012). *A theory of fields* Oxford University Press.
- Gaudin, S. (2011). *AMD moves away from intel rivalry, rethinks course*. Retrieved from <https://www.computerworld.com/article/2499598/computer-processors/amd-moves-away-from-intel-rivalry--rethinks-course.html>
- Gioia, D. A. (2008). *Organizational identity as an emerging perennial domain*. *Sage Handbook of New Approaches in Management and Organization*, , 63-65.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harvard University Press.
- Granovetter, M. (1985). *Economic action and social structure: The problem of embeddedness*. *American Journal of Sociology*, 91(3), 481-510.
- Greene, D., O'Callaghan, D., & Cunningham, P. (2014). *How many topics? stability analysis for topic models*. Retrieved from <https://arxiv.org/abs/1404.4606>

- Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467-483. doi:10.1016/j.tourman.2016.09.009
- The handbook of economic sociology* (1995). . Princeton, NJ, United States: Princeton University press.
- Harvey, D. (2007). *A brief history of neoliberalism* Oxford University Press, USA.
- Hatten, K. J., & Hatten, M. L. (1987). Strategic groups, asymmetrical mobility barriers and contestability. *Strategic Management Journal*, 8(4), 329-342.
- He, H., & Baruch, Y. (2009). Transforming organizational identity under institutional change. *Journal of Organizational Change Management*, 22(6), 575-599. doi:10.1108/09534810910997014
- Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, 332(6025), 60-65.
- Hsu, G., & Hannan, M. T. (2005). Identities, genres, and organizational forms. *Organization Science*, 16(5), 474–490. doi:10.1287/orsc.1050.0151
- Idowu, S. O., & Towler, B. A. (2004). A comparative study of the contents of corporate social responsibility reports of UK companies. *Management of Environmental Quality: An International Journal*, 15(4), 420-437.
- Jacobi, C., van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 4(1), 89-106.
- Joas, H. (1996). *The creativity of action* University of Chicago Press.
- Jung, J. (2015). Shareholder value and workforce downsizing, 1981–2006. *Social Forces*, 93(4), 1335-1368. Retrieved from <https://muse.jhu.edu/article/581904>

- Kobylecky, L. (2016). *AMD vs. intel: Tech's oldest rivalry explained*. Retrieved from <https://www.pastemagazine.com/articles/2016/07/amd-vs-intel-the-truth-behind-techs-oldest-compute.html>
- Koltcov, S., Koltsova, O., & Nikolenko, S. (2014). *Latent dirichlet allocation: Stability and applications to studies of user-generated content.*, 161-165.
doi:10.1145/2615569.2615680
- Koltsova, O., & Koltcov, S. (2013). *Mapping the public agenda with topic modeling: The case of the russian livejournal*. *Policy & Internet*, 5(2), 207-227.
doi:10.1002/1944-2866.poi331
- Lazonick, W., & O'sullivan, M. (2000). *Maximizing shareholder value: A new ideology for corporate governance*. *Economy and Society*, 29(1), 13-35.
- Lievens, F., Van Hove, G., & Anseel, F. (2007). *Organizational identity and employer image: Towards a unifying framework**. *British Journal of Management*, 18, S59.
doi:10.1111/j.1467-8551.2007.00525.x
- LIN JIANG, JUSTIN TAN, & MARIE THURSBY. (2011). *Incumbent firm invention in emerging fields: Evidence from the semiconductor industry*. *Strategic Management Journal*, 32(1), 55-75. doi:10.1002/smj.866
- Livengood, R. S., & Reger, R. K. (2010). *That's our turf! identity domains and competitive dynamics*. *Academy of Management Review*, 35(1), 48-66.
- Luhtakallio, E., & Yi-Anttila, T. (2011). *Julkisen oikeuttamisen analyysi sosiologisena tutkimusmenetelmn*. *Sociologia*,
- Macher, J. T., Mowery, D. C., & Di Minin, A. (2007). *The "non-globalization" of innovation in the semiconductor industry*. *California Management Review*, 50(1), 217-242.

- Miller, I. M. (2013). *Rebellion, crime and violence in qing china, 1722–1911: A topic modeling approach*. *Poetics*, 41(6), 626–649. doi:10.1016/j.poetic.2013.06.005
- Mohr, J. W., & Bogdanov, P. (2013). *Introduction—Topic models: What they are and why they matter*. *Poetics*, 41(6), 545–569. doi:10.1016/j.poetic.2013.10.001
- Mtzel, S. (2015). *Facing big data: Making sociology relevant*. *Big Data & Society*, 2(2), 2053951715599179.
- Munksgaard, R., & Demant, J. (2016). *Mixing politics and crime – the prevalence and decline of political discourse on the cryptomarket*. *International Journal of Drug Policy*, 35, 77–83. doi:10.1016/j.drugpo.2016.04.021
- Peteraf, M., & Shanley, M. (1997). *Getting to know you: A theory of strategic group identity*. *Strategic Management Journal*, , 165-186.
- Plos, L., Hannan, M. T., & Carroll, G. R. (2002). *Foundations of a theory of social forms*. *Industrial and Corporate Change*, 11(1), 85-115.
- Podolny, J. M., Stuart, T. E., & Hannan, M. T. (1996). *Networks, knowledge, and niches: Competition in the worldwide semiconductor industry, 1984-1991*. *American Journal of Sociology*, 102(3), 659-689.
- Polanyi, K. (1957). *The great transformation:(The political and economic origin of our time)* Beacon Press.
- Poole, R. (2016). *Good times, bad times: A keyword analysis of letters to shareholders of two fortune 500 banking institutions*. *International Journal of Business Communication*, 53(1), 55-73.
- Porter, M. F. (1980). *An algorithm for suffix stripping*. *Program*, 14(3), 130-137.

- Porter, M. E. (1980). *Competitive strategy: Techniques for analyzing industries and competition*. New York, 300
- Ragin, C. C., & Amoroso, L. M. (2010). *Constructing social research: The unity and diversity of method* Pine Forge Press.
- Rao, H., Davis, G. F., & Ward, A. (2000). *Embeddedness, social identity and mobility: Why firms leave the NASDAQ and join the new york stock exchange*. *Administrative Science Quarterly*, 45(2), 268-292. doi:10.2307/2667072
- Rekom, J. v., & Riel, C B M van. (2000). *Operational measures of organizational identity: A review of existing methods*. *Corporate Reputation Review*, 3(4), 334-350.
doi:10.1057/palgrave.crr.1540124
- Rutherford, B. A. (2005). *Genre analysis of corporate annual report narratives: A corpus linguistics-based approach*. *The Journal of Business Communication* (1973), 42(4), 349-378.
- Sievert, C., & Shirley, K. E. (2014). *LDavis: A method for visualizing and interpreting topics*. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63-70.
- Simon, H. A. (1991). *Bounded rationality and organizational learning*. *Organization Science*, 2(1), 125-134.
- Singer, G. (2012). *The rise and fall of AMD*.
- Teece, D. J. (1986). *Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy*. *Research Policy*, 15(6), 285-305.
doi:10.1016/0048-7333(86)90027-2

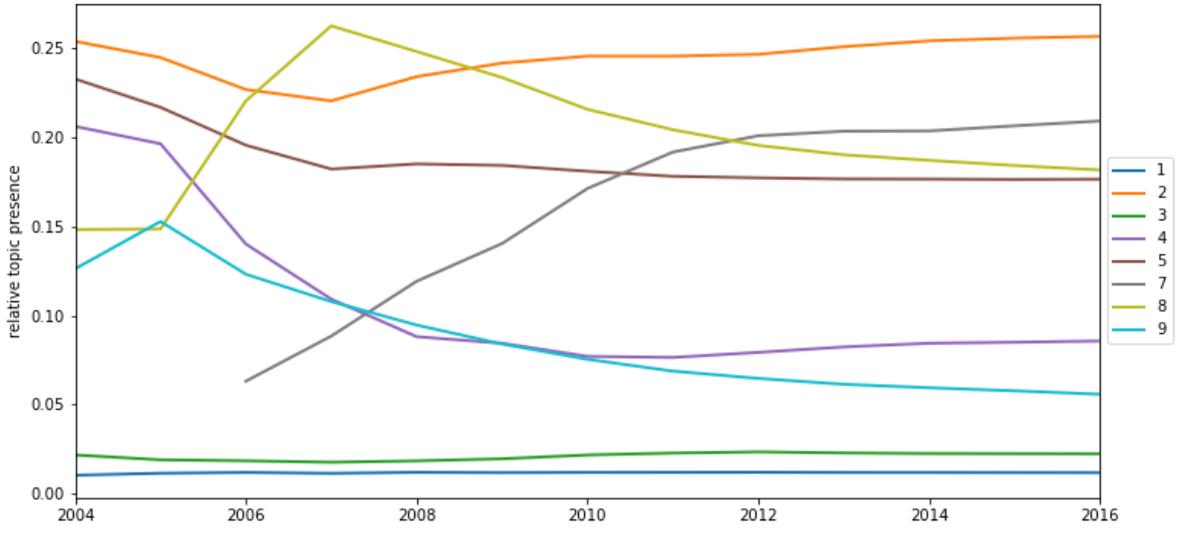
- Tripsas, M. (2009). Technology, identity, and inertia through the lens of "The digital photography company". Organization Science, 20(2), 441-460.*
doi:10.1287/orsc.1080.0419
- Ttt, P. (2004a). Syvllist ja pinnallista: Teoria, empiria ja kausaalisuus sosiaalitutkimuksessa Vastapaino.*
- Ttt, P. (2004b). Syvllist ja pinnallista: Teoria, empiria ja kausaalisuus sosiaalitutkimuksessa Vastapaino.*
- Ubilla, A. (2017, -02-13T16:34:29-05:00). AMD vs intel – A lesson from information technology school. Retrieved from*
<https://iticollege.edu/amd-vs-intel-a-lesson-from-information-technology-school/>
- Vandemaele, S. N., Vergauwen, P. G. M. C., & Smits, A. J. (2005). Intellectual capital disclosure in the netherlands, sweden and the UK. Journal of Intellectual Capital, 6(3), 417-426. doi:10.1108/14691930510611148*
- Whetten, D. A., & Mackey, A. (2002). A social actor conception of organizational identity and its implications for the study of organizational reputation. Business & Society, 41(4), 393-414.*
- White, H. C. (1981). Where do markets come from? American Journal of Sociology, 87(3), 517-547.*
- Wiedemann, G. (2013). Opening up to big data: Computer-assisted analysis of textual data in social sciences. Historical Social Research/Historische Sozialforschung, , 332-357.*
- Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. Proceedings of the 22nd International Conference on World Wide Web - WWW '13, doi:10.1145/2488388.2488514*

6.2 - Appendix

6.2.1 - Full LDA model outputs

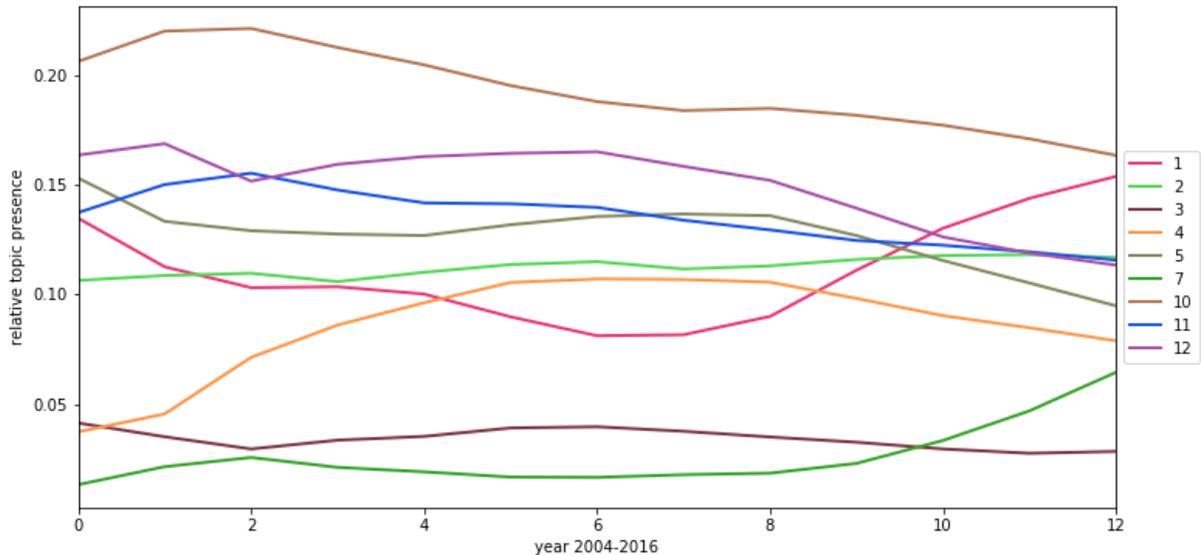
AMD

	1	2	3	4	5	7	8	9
1	brand	market	product	design	intel	graphic	product	includ
2	market	intel	custom	market	microprocessor	chipset	industri	memori
3	support	product	exclus	base	system	market	price	technolog
4	expenditur	microprocessor	potenti	includ	comput	integr	competit	competitor
5	advertis	aggress	market	processor	standard	competitor	introduc	flash
6	posit	price	strategi	offer	market	discret	result	market
7	intel	research	current	competitor	domin	intel	compet	share
8	custom	greater	bundl	compani	manufactur	manufactur	perform	price
9	consum	develop	payment	develop	motherboard	could	abil	busi
10	electron	financi	merchandis	solut	chipset	microprocessor	technolog	action
11	competit	resourc	channel	support	includ	domin	increas	compet
12	whose	lower	partner	abil	control	includ	time	alloc
13	microprocessor	competit	small	licens	industri	sell	processor	practic
14	typic	technolog	mainstream	embed	compani	offer	includ	strategi
15	increas	manufactur	build	parti	graphic	perform	standard	intel
16	materi	domin	super	third	platform	competit	vendor	limit
17	advers	custom	core	provid	posit	nvidia	electron	nvidia
18	effect	substanti	mobil	product	compon	supplier	capabl	rebat
19	affect	advers	nvidia	competit	facto	mainstream	corpor	primarili
20	product	compani	electron	power	logic	user	develop	consol
21	includ	posit	result	intel	chip	use	exist	design
22	remain	result	requir	corpor	depend	bundl	competitor	spansion
23	silicon	expect	futur	architectur	influenc	particularli	brand	margin
24	process	sale	whose	tablet	addit	proprietary	improv	corpor
25	matrox	averag	limit	microprocessor	result	interfac	signific	categori
26	visual	margin	altern	factor	program	processor	introduc	product
27	solut	continu	workstat	server	interfac	place	softwar	imagin
28	result	affect	netbook	compet	distribut	disadvantag	semiconductor	storang
29	exist	corpor	announc	depend	drive	inform	cost	oper
30	lesser	princip	rel	princip	access	give	avail	enter



Intel

	1	2	3	4	5	7	10	11	12
1	product	product	product	perform	architectur	product	facil	standard	product
2	segment	develop	commun	abil	offer	technolog	manufactur	specif	chipset
3	competitor	technolog	network	product	microprocessor	secur	compani	industri	compet
4	market	compet	oper	provid	corpor	market	competitor	particip	memori
5	compani	competit	includ	price	processor	segment	assembl	product	compani
6	competit	chang	offer	competit	product	offer	foundri	adopt	flash
7	busi	abil	processor	solut	compet	consum	network	organ	market
8	compet	price	design	compet	competitor	center	control	initi	graphic
9	offer	cost	multipl	offer	busi	provid	servic	market	microprocessor
10	comput	custom	connect	improv	softwar	competit	patent	intel	segment
11	commun	factor	compet	energi	rival	compani	licens	softwar	incorpor
12	softwar	reduc	broadcom	effici	develop	busi	product	develop	nvidia
13	devic	introduc	wire	platform	base	process	subcontractor	advantag	busi
14	larg	industri	wireless	comput	market	digit	parti	system	compon
15	marketplac	process	current	variou	devic	gener	third	platform	model
16	increas	cycl	compani	compon	compat	protect	design	comput	perform
17	emerg	short	storag	support	server	solut	capit	usag	corpor
18	numer	requir	infrastructur	recognit	compani	use	time	offer	technolog
19	multipl	market	system	integr	design	comput	semiconductor	oper	offer
20	establish	segment	manufactur	believ	advanc	growth	qualiti	compet	includ
21	industri	standard	parti	avail	machin	area	enabl	abil	improv
22	entrant	architectur	third	brand	depend	servic	direct	continu	demand
23	includ	semiconductor	athero	featur	includ	featur	invest	power	competit
24	supplier	typic	qualcomm	innov	success	believ	requir	technolog	desktop
25	provid	evolv	invest	includ	microsystem	signific	advantag	latest	semiconductor
26	platform	smaller	wimax	depend	micro	transit	corpor	content	integr
27	custom	experi	semiconductor	custom	intern	major	addit	associ	samsung
28	compon	competitor	smaller	worldwid	competit	cloud	offer	model	system
29	continu	increas	optim	power	continu	develop	need	competitor	support
30	architectur	platform	fpga	secur	embed	perform	factor	commun	current



6.2.2 - The code for LDA

```

from nltk.tokenize import RegexpTokenizer
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from gensim import corpora, models
import gensim
import pyLDAvis.gensim
import datetime
from os import chdir
import re
from collections import Counter
import pandas as pd
import glob
import matplotlib.pyplot as plt
import matplotlib
import numpy
from nltk.stem.wordnet import WordNetLemmatizer

en_stop = list(stopwords.words('english'))
tokenizer = RegexpTokenizer(r'\w+')
p_stemmer = PorterStemmer()
lemma = WordNetLemmatizer()

def corpusopener(choice):
    years = range(2000-2017)
    for item in years:
        list.append(en_stop, str(item))

    if choice == "paragraph":
        for filename in glob.glob('*.*txt'):
            with open(filename, "r") as f:
                for line in f:
                    if len(line) > 40:
                        list.append(doc_set, line)

    return doc_set

def corpushandler(doc_set):
    texts = []

```

```

for i in doc_set:
    # clean and tokenize document string
    raw = i.lower()
    tokens = tokenizer.tokenize(raw)
    stopped_tokens = [i for i in tokens if not i in en_stop]
    stopped_tokens2 = [i for i in stopped_tokens if len(i) > 4]
    normalized = [lemma.lemmatize(word) for word in stopped_tokens2]
    stemmed_tokens = [p_stemmer.stem(i) for i in normalized]
    texts.append(stemmed_tokens)
all_tokens = sum(texts, [])
c = Counter(all_tokens)
ten_most_common = c.most_common(15)
ten_most_common = dict(ten_most_common)

keys = ten_most_common.keys()

tokens_once = set(word for word in set(all_tokens) if all_tokens.count(word) < 3)
# remove too common words
#remove too few words
texts = [[word for word in text if word not in tokens_once] for text in texts]
all_tokens2 = sum(texts, [])
print(len(all_tokens2))
# turn our tokenized documents into a id <-> term dictionary
dictionary = corpora.Dictionary(texts)
# convert tokenized documents into a document-term matrix
corpus = [dictionary.doc2bow(text) for text in texts]
return corpus, dictionary

def ldamrunner(corpus, dictionary, num, pass_num):
    # generate LDA model
    ldamodel = gensim.models.ldamodel.LdaModel(corpus, num_topics=num, id2word =
dictionary, passes=pass_num)
    sparse_matrix = gensim.matutils.corpus2csc(corpus)
    return ldamodel, sparse_matrix

def topics_to_dataframe(num_of_words):
    docTopicProbMat = ldamodel[corpus]
    K = ldamodel.num_topics
    topicWordProbMat = ldamodel.print_topics(K, num_words=num_of_words)
    x = 1
    for i in topicWordProbMat:
        # print (i[0], i[1])
        list_of_words = i[1].split("+")
        wordlist = []
        for i in list_of_words:
            word1 = " ".join(re.findall("[a-zA-Z]+", i))

            list.append(wordlist, word1)
        if x == 1:
            tehdata = pd.DataFrame({str(x) : wordlist})
        if x > 1:
            tempdata = pd.DataFrame({str(x) : wordlist})

            tehdata = pd.DataFrame.join(tehdata, tempdata)
        x += 1
    return tehdata

#input by paragraph
choice = "paragraph"

#choose the number of topics
num = 12

#choose number of iterations

```

```
pass_num = 300

#possible firms: AMD, INTEL
firm = "INTEL"

#directory of files
chdir('c:/datasets/729/ALL/INTEL/')

# call functions
doc_set = []
doc_set = corpusopener(choice)
corpus, dictionary = corpushandler(doc_set)
ldamodel, sparse_matrix = ldarunner(corpus, dictionary, num, pass_num)

#save topics to dataframe
topic_df = topics_to_dataframe(30)

# visualize results
pyLDAvis.enable_notebook()
pyLDAvis.gensim.prepare(ldamodel, hojo, dictionary)
```

6.2.3 - Links to the sec database

<https://www.sec.gov/edgar/searchedgar/companysearch.html>

(link to a sample report (Intel 2015)):

<https://www.sec.gov/Archives/edgar/data/50863/000005086316000105/a10kdocument12262015q4.htm>