A Work Project, presented as part of the requirements for the Award of a Masters Degree in Management from the NOVA School of Business and Economics.

Digital Acceleration Index:
Digital Maturity trends in Portugal

Daniel Dormal André

A Project carried out on the Masters in Management Program, under the supervision of:

Professor João Duarte

May, 2021

Abstract

The importance of having the digital means to adapt and survive digital disruption is a ubiquitous factor for firms in the 21st century. The Digital Acceleration Index (DAI) is a Digital Maturity Model developed by Boston Consulting Group (BCG). This project follows its first edition in Portugal, analyzing the results and providing a framework for analyzing and mapping out digital trends through text analysis. A second sister-thesis further explores the relationship between Digital Maturity and gross salary through the lens of a multiple regression model.

Keywords: Digital Maturity, Digital Transformation, Maturity Model

Acknowledgments:

This work had the support of Boston Consulting Group, Nova Junior Consulting and the Digital Experience Lab from Nova School of Business and Economics. A further acknowledgement is made to the invaluable contribution of my thesis partner and co-author João Azevedo, without whom this project would not have been possible.

Table of Contents

List of Figures

List of Tables

1 Introduction

The digital world has become increasingly interwoven with our own physical reality. The distinction between physical and digital has become blurred, with companies aiming to create end-to-end experiences, products and services that may start digitally and end analogically. Disruption has become ubiquitous and a cause of concern for consumers, firms and even entire nations. This shift in paradigm has seen new concepts and terminology, namely Digital Maturity, appear during the last decade, alongside attempts at measuring it based on broad ranging digital frameworks. Companies will find benefits in going digital and compete for ”digital supremacy” (Bloom and Pierri, 2018), in other words, companies who set digital transformation as a priority and see it through will reap the economic benefits and potentially be able to rise far above the competition. In fact, recent studies have highlighted the importance of going digital in order to enhance productivity (Gal et al., 2019).

Attempts at measuring Digital Maturity began in the early 2010s with domain-specific practitioner-based models based on empiric observation. Over the past decade, we have seen several digital maturity frameworks pop-up, both industry-specific and nationwide. However, most of these models fail to account for an important variable: the evolution of digital maturity over time.

The ultimate goal of this study is firstly to portray the state of Digital Maturity in Portugal; second, to see how that maturity influences the existing firms and thirdly; what digital transformation trends can be observed from Portuguese managers' inputs. Through the 2020 edition of the Digital Acceleration Index, a Digital Maturity Model developed by the Boston Consulting Group, and the first of its kind in Portugal, 1044 firms from several different industries are surveyed regarding 36 different digital topics (hereafter referred to as ”dimensions”). The results are first analyzed with an exploratory mindset and then by economic sector and firm size. In what is ultimately a sparsely inhabited and still nascent field, this is the first step in accounting and studying the evolution for Portuguese firms' Digital Maturity over the next decade.

In order to further understand the trends surrounding the topic in Portugal, a qualitative framework has been created for the analysis of keywords in Digital Maturity through word-clouds (also called ”tag clouds”) and through network structure analysis.

This project starts by going through the existing literature regarding digitalization, namely, some of the concepts and challenges in defining and measuring digital maturity and what has been researched about the Portuguese economy. This project also goes into detail regarding the methodology behind the DAI survey data collection and data curation. Results from the survey and subsequent analysis are then presented and interpreted, with those being contextualized in the existing literature and limitations found in the study. Suggestions for future developments are also made.

This thesis was a collaborative project and the relationship between Digital Maturity and human capital can be read in the work of its co-author (Azevedo, 2021).

2 Literature Review

2.1 Defining Digital Maturity

When discussing the transition from the physical to the digital world, technical terms like Digitalization and Digitization are frequently used interchangeably. However, some scholars have made an effort to define these topics in a more precise and nuanced way. Brennen and Kreiss (2016) base their definition on the Oxford English Dictionary's entry, stating that digitization consists of any process of converting analog information into ”digital bits of 1s and 0s” while Digitalization refers to the general adoption of technology in many domains of society. Thusly, from a management perspective, one can see an example of digitization when a firm migrates their transaction records previously kept on paper to digital. Digitalization, on the other hand, can be seen in new product or service offers or even broader changes to a firm's Business Model in order to adapt and adopt new available technologies and improve performance (Racinger et al, 2018).

The definition of digital maturity is often blurred by differences in perspectives (Aslanova and Kulichkina, 2020). At its core, it aims to represent the status of a firm at a given time, not only through its current processes and adoption of new technology but also in its flexibility and preparation in adapting to future changes. Often disruptive, they represent shifts in paradigm as seen through so-called ”soft” digital attributes - attributes such as corporate culture, leadership and governance, among several others. Digital Maturity is, therefore, a way of assessing the progress of a given company in its digital transformation process (Teichert, 2019).

Mapping out the path undergone by firms is an imprecise challenge. Teichert's (2019) systematic review of more than 20 Digital Maturity Models (DMMs) reveals that there is no consensus regarding the nature and definition of digital transformation. Furthermore, variations occur at many different levels. They appear between practitioners and academics, between general and domain-specific approaches and also based on the underlying digital framework and roadmap.

One of the most poignant differences when reviewing DMMs are based on whether a model is Academic or Practitioner-based (Felch et al., 2019). This is true both in matters of substance - science-based models are more descriptive in nature when compared to the prescriptive nature of practitioner DMMs.

The first models considered by Teichert (2019) were all practitioner-based, with academic DMMs being published only years after. In light of this, academic models built on top of the groundwork defined by practitioner-based models. Either one is based on empirical data/observation.

Teichert (2019) also gives us the definition of dimension as a “specific, measurable and independent component which reflects a major, fundamental and distinct aspect of digital maturity” and in his systematic review of 22 digital maturity models, he finds that all of them assess between 4 and 9 dimensions, of which the most common are: “Digital Culture”, “Technology”, “Operations & Processes” and “Digital Strategy”. On the other side, there is a distinct lack of attention given to the “Business Model” dimension, with only 3 out of 22 models assessing it.

It is also worth mentioning that, despite culture-related dimensions being present throughout most models, there is an inconsistency as to what culture attributes should be assessed and if the information collected is reliable. Buvat et al. (2017) expose the huge disparity between what the senior executives think of their firm’s digital culture compared to what the employees think. For example, in France, 45% of surveyed leaderships thought that there was a strong digital culture in their organization contrasted with the 3% of employees in agreement.

Another point of contention is the lens through which Digital Maturity progress can be assessed. Some studies view Digital Maturity as a linear process that undergoes a set number of stages, as seen with Berghaus and Back (2016) and Leyh et al. (2016), among others. In practical terms, what these studies seem to have in common is an attempt at mapping a firm’s unique configuration to a defined stage of Maturity, measured across the variables within the scope of the model. The stages vary between models, the authors above state 5 and 4 possible states respectively while Strategy&’s “Measuring industry digitization” (Friedrich et al., 2011) splits firms into 3 stages (laggards, midfield and leaders). In spite of the nuance imbued in each of the models above, these stages convey approximately the same in regards to maturity. In contrast, non-linear models view Digital Maturity as a fluid state, resulting from a mix of different qualities that don’t necessarily progress in a fixed order. Remane et al. (2017) further expands on this notion by suggesting a framework that takes a firm's Digital Impact (the effect of digital transformation) and Digital Readiness (the preparedness to tackle disruption and undergo further digital transformation) into account, offering 4 possible configurations conveyed through a 2x2 matrix. This view offers a dynamic configuration, but has the disadvantage of complicating the process of ranking different firms.

Conversely, regarding the measurement of the maturity level, most studies seem to mimic the same method, relying on surveys and self-assessment descriptions to collect information about the respective firms (Barata and Cunha, 2017). Most of the bigger surveys are conducted periodically by major consulting groups such as PWC's Industry 4.0 report, KPMG's Digital Readiness Assessment or Deloitte's Digital Maturity Model.

This project focuses on the very relevant topic of enterprise digital maturity, with a focus on Portugal's firms. It is therefore important to explore the thematics surrounding Digital Maturity and what it means for a firm to achieve ”maturity” in the digital spectrum.

While this project is focused on the digital maturity of enterprises operating in Portugal, it is relevant to investigate what has been researched regarding the impact and challenges of "going digital" in the business world and how these insights can apply to the Portuguese economic landscape.

As previously mentioned, most of the major digital maturity surveys and analysis are done by Consulting groups and more recently there have been some academic papers focusing on the subject. However, Portugal seems to have a research gap in DMMs, having no major practitioner or academic analysis of the economy's Digital maturity . In fact, most of the existing information about the Portuguese economy's digital level is from studies done by European Union and government entities.

Such is the case of the Digital Economy and Society Index (DESI), a yearly report on the degree of digitalization of each of the 27 member states plus the UK, that focuses on 5 dimensions (each with a different score): "Connectivity", "Human Capital", "Use of Internet services", "Integration of Digital technologies" and "Digital public services". Results from 2020 showed that despite being above the EU average in Connectivity and Digital public services, Portugal is quite behind the curve when it comes to (a) Human Capital (21st out of 28), mostly due to having a large percentage of people who have no digital skills (26%) and lacking a significant percentage of ICT graduates (only 1.9% of all college graduates) and (b) Use of Internet Services (24th out of 28) by having a large percentage of the population that has never used the internet (22% vs. EU average of 9%).

Additionally, the lack of knowledge can also be observed in a managerial sense, as many Portuguese managers don't know how and where to implement digital improvements in their firms and lack awareness of the respective potential benefits which, in turn, leads to low willingness to invest and a low adhesion to existing public financing programs (European Investment Bank. 2019).

Effectively, this digital literacy gap in the Portuguese economy seems to be in line with the findings from other reports that also point this out as a decisive factor for the country's divergence from the EU average.

2.2 Impact of Digital Maturity on firm outcome

The Covid-19 Pandemic and economic duress stemming from national lockdowns brought to the limelight the importance of being up to date with all the technological tools, highlighting the severe shortcomings in many companies and industries that failed to catch up with new advancements in Information and Communications Technologies (Baig et al., 2020; Fletcher and Griffiths, 2020). According to Summa (2016), a firm's adaptability and reaction time to sudden change might be make-or-break factors in its success. Naturally, this agility is often tied to a company's digital prowess.

Effectively, surveys and data from the past year seem to show how companies have been pushed by the pandemic to transform their businesses. A majority of these firms also view this shift as long-lasting transformation instead of a temporary adjustment.

At first glance, the path to digital maturity seems like an obvious choice for any company looking to grow. There is extensive research detailing the benefits of going digital on channels such as Marketing (Kannan and Li, 2016), Big Data management (Wamba et al., 2014), Industrial processes (Dalenogare et al., 2018) and Supply Chain management (Büyüközkan and

Göçer, 2018). Grebe et al. (2018), who followed the performance of a group of firms over 3 years, also point out that digitally mature companies outperform their peers in metrics such as Time to Market and Cost efficiency.

However, Fletcher and Griffiths (2020) warn against introducing new technology as a knee-jerk reaction before doing their due diligence. When lacking required infrastructure and readiness, these could instead pose a threat to the firm, as is the case of cybersecurity. Recently, firms without videoconferencing capabilities resorted to consumer grade messaging applications as a solution. While seemingly innocent, these actually represent a threat to company data and are potentially non-compliant with European Data Protection Laws. Furthermore, they also reflect poor digital governance.

Therefore, it is not correct to assume that an enterprise should just invest blindly in digitalization and expect to have widespread gains. In fact, even under good conditions, not all digital investments net the same return. A model done by McKinsey Digital (Bughin et al., 2017) attempted to measure the average return on investment for different digitalization efforts and found that investing in digital Products and Processes seems to offer a bigger increase in revenue (3.5% and 3.2% respectively) when compared to digitalization attempts in Supply Chain (2.3% increase).

To sum up, digitalization is a complex process that requires more than the mere adoption of new technologies. A company must also have a solid foundation in terms of work methods and organizational agility in order to adapt quickly and have human capital that is skilled and up-to-date with new ICT (Frey and Osborne, 2013). Additionally, Dalenogare et al. (2018) take it a step further and point out that the external environment (here defined as the general level of digital maturity of the surrounding economy) is also a key determinant to the success of digital transformation, namely in developing countries that often lack the digital infrastructure to support state-of-the-art digitalization.

In fact, some researchers argue that the new digital transformation wave has contributed to create a larger gap in firm performance since the companies that are better suited to have productivity benefits from digital improvements, are already the most productive in the first place (Gal et al., 2019). This can create a virtuous cycle for the digital leaders by attracting the highest skilled workers since digital maturity has also been associated with an increase in skilled labor (Bartel et al., 2007) due to the way it helps automate repetitive tasks and create more challenging job positions. That being said, since some studies have pointed to a positive relationship between productivity and salaries (Laskina et al., 2020) one could theorize that digital maturity might have an indirect positive influence on salaries due to productivity gains. Based on the research done on the literature, there doesn't seem to be any major publications answering this question directly apart from a study from 2001 surveying a group of Taiwanese manufacturing firms that found significant results confirming a positive relationship but with the caveat that results might differ from country to country (Liu et al., 2001).

3 Methodology

This thesis relies mainly on the results from a survey sent out to Portuguese companies, collected by Nova School of Business and Economics and the Institute of Marketing Research (IMR). In the sister-thesis of Azevedo (2021), a second dataset regarding a nation-wide census of the Portuguese Workforce data for 2017 provided by the Instituto Nacional de Estatística (INE) utilized.

The survey data is comprised by rating statements and optional open-ended questions in regards to those, therefore providing a plethora of contextual and text data in addition to the numerical scores used to calculate the Digital Acceleration Index.

3.1 Digital Acceleration Index (DAI)

The Digital Acceleration Index is a metric used to evaluate a firm's digital maturity over time. Unlike other metrics, which measure the state of digital maturity at a fixed point in time, the goal of the Digital Acceleration Index applied at a national scale (DAI@Scale) is to construct a comprehensive image of an economy's digital maturity as well as its evolution.

It is a practitioner-based model created by the Boston Consulting Group's (BCG) based on their holistic digital framework. The first iteration of the DAI@Scale took place in Singapore 2019, and similar editions of this survey are being made in other countries. In the interest of measuring digital maturity over time, this survey should be the first of a systematic effort to record and document the evolution of Portuguese companies' digital maturity every couple of years.

3.1.1 Definition

Like most Digital Maturity Models, the Digital Acceleration Index attempts to measure all facets of a company end-to-end. It contemplates 36 different processes or dimensions divided along 6 major blocks, as seen in Figure 1. The formula for the DAI itself corresponds to the arithmetic average of these dimensional scores converted to a scale from 0 to 100:

$$DAI = \frac{1}{n}\sum_{i=1}^{n} a_i * 100/4$$

Where n is the number of dimensions (34, 35 or 36) applicable - the Manufacturing and Shared Service Centers are firm specific - and a is the score from 0 to 4, the former used in cases where the respondents were unable to quantify the dimension.

An adjusted version (DAI Adjusted) was developed on the basis that it portrays a more accurate self-image from the firms' survey answers, due to a particularity in large parts of the data collected. The DAI Adjusted is calculated through the following adjustment, with "empty dim" representing "empty dimensions" or "don't know" replies in dimensions:

$$DAI = \frac{1}{n - empty\ dim}\sum_{i=1}^{n} a_i * 100/4$$

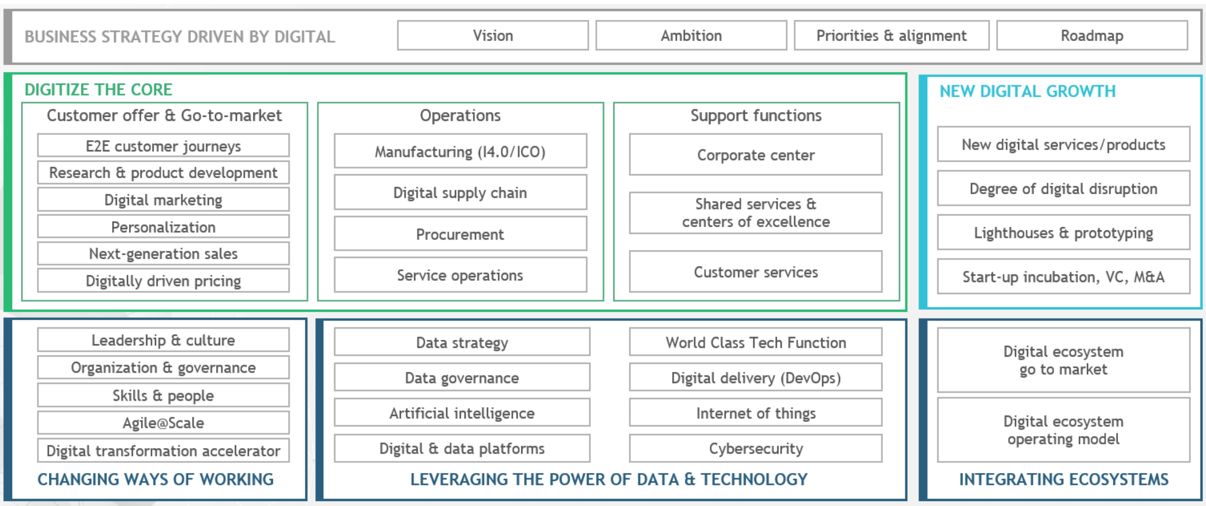

Figure 1: Boston Consulting Group's Digital Holistic Framework

The first block, ”Business Strategy Driven by Digital”, starts by assessing whether a company’s overall business strategy has Digital at the forefront of it or whether the company at least admits the importance of having a clearly defined goal and respective roadmap communicated across its entirety.

The second and largest block is ”Digitize the Core”, which encompasses all the ”essential” business sectors and is divided into 3 sub-blocks: ”Customer Offer & Go-to-market”, which assesses the degree to which a company is using digital solutions to enhance their customer journey, marketing and sales; ”Operations” which focuses on the presence of Industry 4.0 technologies in manufacturing and supply chain and ”Support Functions” which measures the rate of Robotic Process Automation (RPA) or Artificial Intelligence (AI) usage in repetitive and high-volume activities such as customer service, accounting, and payroll.

The third block, ”New Digital Growth” is dedicated to Digital Innovation. Namely, it looks into how the firm uses the digital space to develop new products and services and how that development process takes place. Additionally, it also inquires about inorganic development through M&As and start-up incubation/investment.

”Changing Ways of Working” is part of a broader set of blocks dubbed the ”enablers”. It focuses both on the overarching work style as well as the hiring and retention of digital talent. Furthermore, it also encompasses dimensions related to company culture and to whether Digital is cascaded top-down from the board of directors to the rest of the workforce.

The fifth block is ”Leveraging the power of Data & Technology”. Comprised of 8 dimensions, this block gives us an overview of if and how the company uses data as a driver for growth. Ranging from overall data strategy & governance (e.g. does the company have a dedicated Chief Data Officer) to more nuanced areas such as Internet of Things (are there dedicated teams for development of use cases), Artificial Intelligence (is AI integrated in the decision-making process) and Cybersecurity (is it deployed on an ad-hoc basis or is the company taking preemptive measures and actively looking into its own vulnerabilities).

The last block of the enablers and the framework itself is ”Integrating Ecosystems”, which focuses on the existence (or lack thereof) of Digital Ecosystems between the responding firm and its business partners.

3.1.2 Survey/Data Collection

For each of the 36 dimensions, respondents were asked to rate their company's Digital Maturity level for both their current state (2020) and their expected future state (2022) in an ordinal scale from 1 ("Digital Starter") to 4 ("Digital Leader"), with two intermediate options ("Digital Literate" and "Digital Performer"). Each of these dimensions were accompanied by a custom description specifying several practices or standards expected from a company on each of those stages. The respondent was also asked to rate the importance of the question for their organization from a scale of 1 (low importance). For all these questions, there was also a 5th option (”Don't know”) that would allow the respondent to skip ahead.

Respondents were also encouraged to provide further detail on each of their answers through open-ended questions; one inquiring about the respondent's chosen score for its current state and another asking why the company had yet to arrive at the expected future stage in 2022. These questions totaled 180, or five questions for each of the 36 dimensions.

The complete survey also included three initial questions pertaining to the responding company's sector, geographical location and number of workers, two yes or no questions regarding the existence of logistics services or shared services center within the firm and 7 additional context specific questions unique to Portugal and its framing within Europe. This included topics such as employee development, financing options and an open-ended question regarding major challenges faced in the path to digitalization, bringing the total tally of questions in the survey to 192.

As previously mentioned, survey results were collected by two parties: Nova School of Business and Economics' Masters students in collaboration with Nova Junior Consulting and the Institute of Marketing Research (IMR). In the surveys collected directly by the former, each respondent, upon agreeing in principle to participate in the study, was invited to a 30-minute video conference call to be introduced to the goals and objectives of the survey and study. These meetings also served an important second purpose of assuaging participant's concerns regarding privacy and anonymity. Videoconference calls were conducted with at least one person from within the organization - typically the Single Point of Contact (SPoC) - but would often include other stakeholders from within, including cross-disciplinary teams tasked with answering the survey. Respondents would then be given a printable version of the survey for internal use, a link to the survey submission itself and a copy of the presentation deck used in the presentation call. Students would also be available for further follow-ups and direct help in filling out the survey.

3.1.3 Survey Data Cleaning & Outlier Analysis

Unlike the group handled by Nova School of Business and Economics and BCG, most of the firms did not get such rigorous mentoring and support in the process of filling the survey. Therefore, there were a few entries that proved to have little use for any significant analysis, either because many questions were left unanswered or because the survey wasn't answered with the expected due dilligence. Consequently, some data cleaning was necessary to obtain a viable sample.

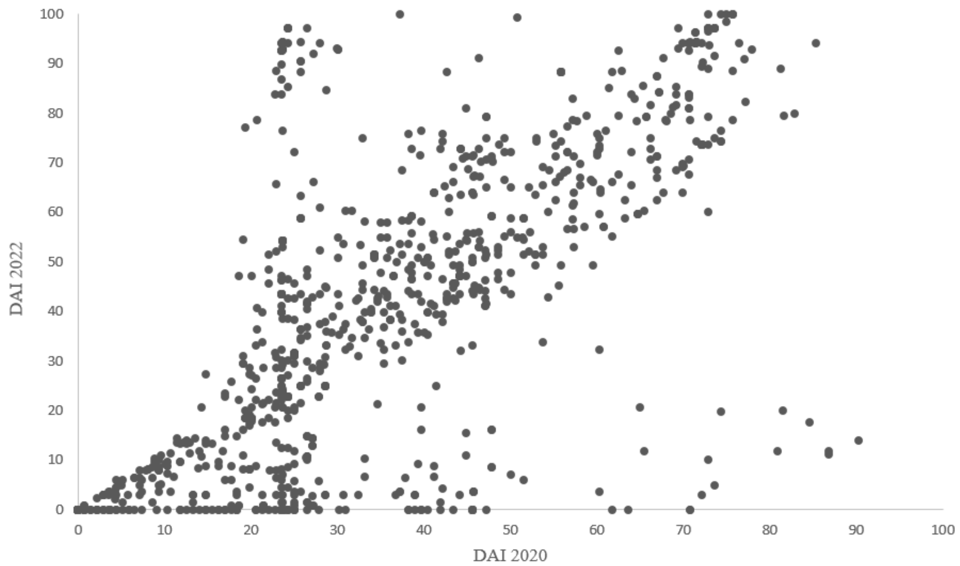

Figure 2: DAI Score 2020 and 2022 expectations

When taking a first glance into the plot above , instantly there were some oddities to be noticed. First of all, since the DAI score expectations for 2022 are on the Y axis and the 2020 DAI scores are on the X axis, all dots below the identity line (Y=X) were potential outliers, since they implied that the responding company expected their digital maturity to decrease over the next 2 years. Upon closer analysis, the reason for this was that many respondents did not fill in an answer for their 2022 expectations, causing a negative distortion of the results. Therefore, the 2022 DAI scores for these companies were discarded, which resulted in a decrease in the 2022 sample from 1044 entries down to 554.

Secondly, we decided to eliminate all entries that failed to answer more than 20 dimensions, out of the total 36. Consequently, our 2020 sample decreased from 1044 to 797 and our 2022 sample decreased to 512.

Lastly, some entries represented highly unrealistic expectations of improvement for such a short period of 2 years, with some respondents aiming to jump as much as 70 DAI points from 2020 to 2022. After looking at the companies selected by Nova which represented some of the portuguese digital leaders and arguably the ones with the biggest capacity to mature digitally in the near future, we decided to cap the DAI gap at 27.5 points, which is the highest expected jump from that pool of companies. Again, this removal only affected 2022 entries, reducing them from 512 to 399.

In conclusion, our final sample was comprised of 797 entries for the DAI in 2020 and 399 entries for the 2022 DAI expectations.

3.2 Natural Language Processing

From the total tally of 192 questions included in the survey, 73 of the questions were presented in an open-ended format; 72 relating to the 36 aforementioned dimensions and one question relating to what companies' felt was their biggest challenge in the process of digitalization. Respondents' accompanied directly by Nova students were also instructed to leave detailed, paragraph-length replies to each of the questions on a basis of importance.

The project's expectation was for this survey to reach at least 1.000 Portuguese companies, which could have resulted in a maximum of 73.000 paragraph length replies left by firms. Although the above scenario was unlikely to materialise, it nevertheless raised the question of how to find and produce valuable insights regarding digital maturity trends from these text-based replies.

Furthermore, the sheer diversity in replies stemming from the 36 dimension-specific questions spread across 6 thematic blocks or the breadth of sectors from the survey respondents would make extracting domain-specific information difficult. As a solution, two range of techniques are proposed to be applied to the survey results; a qualitative approach to sifting the data through wordclouds and a quantitative approach towards analysing the structure of replies through networks.

Both of these techniques fall under the school of Natural Language Processing. The text-based replies would require some additional preprocessing, and those steps go as follow:

Tokenization→Spellchecker→Stopword Removal→Lemmatization

Tokenization is the process of breaking down sentences and text into tokens, or individual words. In this process of finding the tokens within text, special characters, whitespaces, punctuation, etc. are to be removed and not tokenized.

The next step was implemented due to a high likelihood of spelling mistakes to be found in survey replies. A simple dictionary-based spellchecker was passed for each of the tokens to detect errors and suggest corrections based on string similarity with a Portuguese language dictionary.

The next step, stopword removal, is the process of filtering out words that aren't useful or that carry little relevance. This is made with the help of a compiled list of common stopwords in Portuguese. It must be noted, however, that some words will have to be manually removed on a case by case basis, such as the case of a word directly referenced in the question and likely to be used in the reply.

The last step of the process is lemmatization - a morphological analysis and transformation of each in order to obtain a regular or "dictionary form" of a word. This step will convert variants of a verb into the same format, therefore removing some - albeit not all - repetition in visual representation and analysis.

The preprocessed vectors of words will then be visualized through a wordcloud. According to Depaolo and Wilkinson (2014), wordclouds allow a quick extraction of the keywords present in the survey, the which should be relevant to trends in Digital Maturity. Furthermore, we can also extract further value by comparing the totality of our dataset with a smaller subset - in this case, the company surveys that were not accompanied directly by the Nova project team. According to Grebe et al. (2018), the digital champions are likely to set the tone for nation-wide digital maturity, further increasing the interest of contrasting the differing datasets, and understanding how the companies that comprise most of Portuguese firms fare against the bigger players.

According to graph theory, a network or graph provides the mathematical structure used to model relationships between objects and nodes. Tierney (2012) provides a sound basis for the qualitative analysis of networks that, put together with an adjacency matrix, will allow us to study the underlying structure of the survey replies by company, dimension or even by keyword. The following formula will be used to process text data before it is later transformed into an adjacency matrix to be exported into a graph object:

$$TF - IDF = TF(t, d) * IDF(t)$$

Where the Term-Frequency Inverse Document Frequency ($TF - IDF$) is the result of the product of the term frequency - the number of times the term t appears in a survey reply, d - and Inverse Document Frequency (IDF).

The product of the $TF - IDF$ matrix with the transposed version of itself results in a company adjacency matrix, which is then used to create a graph object to be exported and processed with Gephi, an open-source network analysis and visualization software. Since the nodes in the network represent firms, further contextual data (DAI score, firm size, geographical location) can be assigned as node attributes in order to have a better understanding of what influences the underlying structure of this network. As an example, we might find that certain companies cluster together based on a common concern or focus in their replies, and that those companies share a common attribute. Alternatively, the network might also reveal that none of these factors produce viable clusters or segmentation.

4 Results

4.1 Survey Results

The following table contains the summary statistics for the survey results both for the current Digital Maturity state (DAI 2020) and for the expectations for 2022.

As mentioned before, since answers left empty were counted as 0, there was a certain negative distortion of the final scores. Therefore, in order to more precisely see how companies rated their digital maturity, another metric of the index was developed under the name ”Adjusted DAI”. In this case, the only difference when compared to the regular DAI score is that Adjusted DAI is calculated by dividing the sum of answers by the number of dimensions the respondent effectively answered instead of simply dividing by 36.

Table 1: Summary Statistics for the Digital Acceleration Index

	Minimum	Maximum	Mean	SD	n
DAI 2020	16.91	90.28	40.79	18.09	797
Expectations 2022	20	94.28	57.30	19.32	399
Adjusted DAI 2020	25	100	46.43	20.49	797
Adjusted Expectations 2022	25.76	100	65.71	21.02	399

The average DAI Score for the 797 observations was 40.79, which corresponds to a 1.69 in the 1 to 4 scale, placing the Portuguese economic landscape in the ”Digital Literate” category according to BCG’s nomenclature. However, given the diversity of the sample in terms of industries and firm size, it is necessary to breakdown the results in order to paint a correct picture of Digital Maturity in Portugal.

Firstly, looking into the differences in firm size and following the previously defined categorization, in the original sample, the distribution was 99 Big companies, 216 Medium and 729 Small ones, which after the outlier removal were then reduced to 88 Big, 182 Medium and 527 Small companies, therefore keeping a similar distribution to the originally intended one.

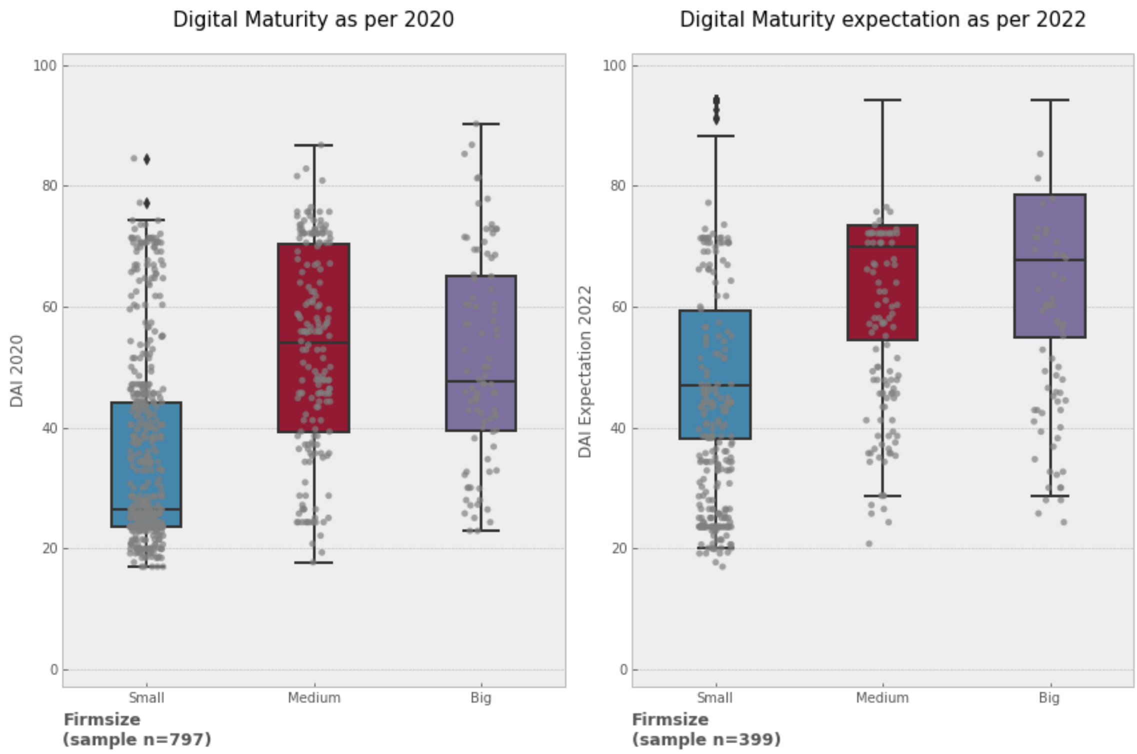

Figure 3: DAI Score distribution by size

In regard to how these unequal sized companies differ in terms of Digital Maturity, the plot above shows that there is a clear difference from the Small firms block compared to the rest of the pack. Presenting an upper quartile only a few points above the lower quartile for the other 2 sizes, both in 2020 and in 2022 expectations, Small firms seem to be at a digital disadvantage to their peers.

On the other hand, Medium and Big corporations do not exhibit such strong contrast. When it comes to 2020, Medium companies presented a bigger median score and upper quartile compared to Big ones, but a very similar average score (52.6 and 51.5, respectively). Conversely, regarding expectations for 2022, Big companies seem to stay ahead in upper quartile and average score (67 points, against 65.5 from Medium sized firms), while still having a slightly smaller median score.

In order to do a sector analysis, the *Classificação Portuguesa de Atividades Económicas* (CAE), an economic activity classification used by the Portuguese National Institute of Statistics (INE), was used to delimit the different industries. In the next table, one can see the sectors that had more than 5 observations and how their average Digital Maturity compares to one another. The original distribution of the 797 firms per sector and size can be seen in Table 1.

Table 2: DAI results by Economic (CAE) Sector

	DAI 2020	DAI 2022	Adj. 2020	Adj. 2022	N
Agriculture & Animal husbandry	29.34	42.61	32.96	47.92	13
Construction	34.64	53.07	39.40	61.06	52
Wholesale & Retail	37.68	56.29	41.93	62.86	171
Hospitality & Tourism	37.88	59.57	43.09	68.54	81
Manufacturing Industries	39.80	53.53	45.14	61.96	195
Healthcare & Welfare	39.93	46.93	45.34	53.71	53
Education	38.92	65.02	45.52	76.44	36
Transportation & Storage	43.50	58.95	48.89	67.11	34
Administrative & Support services	43.26	60.59	49.64	68.45	89
Information and Communication Tech	60.10	70.99	70.58	83.87	58
Financial Services & Insurance	45.38	67.57	53.24	79.44	9

(2022 values are expectations as of 2020) Adj. = Adjusted DAI

At a first glance, it should come as no surprise that the sector with the highest digital maturity is Information and Communication Technologies (ICT) with an average score of 60 coming from a sample of 58 tech companies. This score places them in the Digital Performer bracket [50-75], making it the only sector in Portugal to achieve this feat. Perhaps more impressively, using the Adjusted DAI score, ICT has an average score of 70 which places the sector close to reaching a Digital Leader status. This pole position is further noticeable when looking at the gap from ICT to the sector coming in second: ”Financial Services & Insurance” which pales in comparison by lagging 15 DAI points behind, standing at a score of 43. Interestingly, the remaining sectors do not seem to differ a lot in their scores between each other with only 16 DAI points separating the second place sector from the last place one (Agriculture & Animal Husbandry).

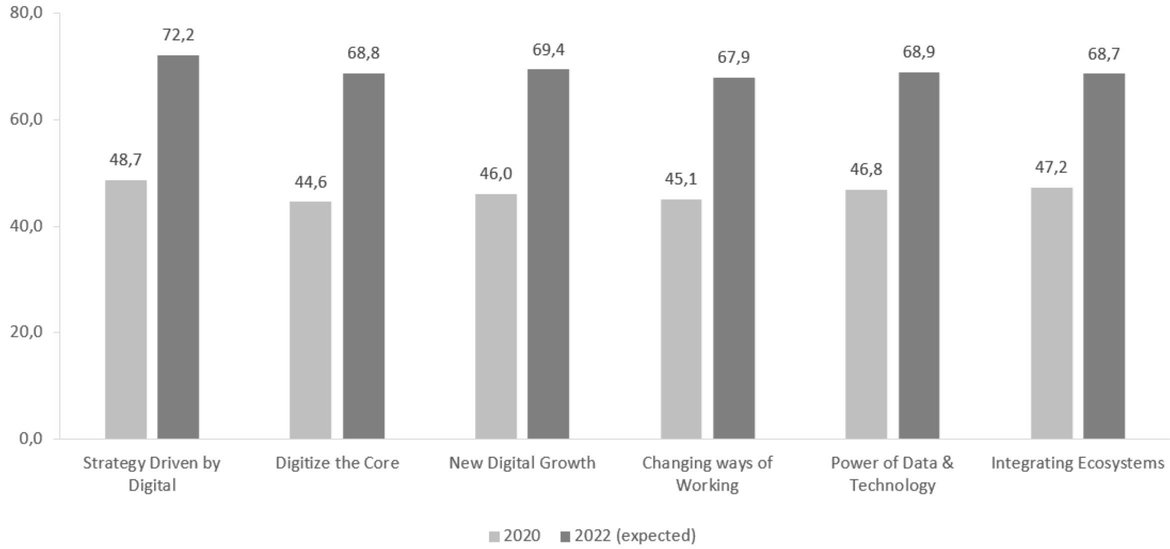

Figure 4: Average Adjusted DAI Scores by block

When it comes to the predictions for maturing digitally in the coming years, the Education industry is the one that stands out the most by expecting an average jump of 26 DAI points, a whole DAI level, from 38.9 (2020) to 65 (2022). On the other end, the Healthcare & Welfare sector predicts only an average 7 DAI point increase in the two year period, making it the less optimistic expectation out of the sample. Interestingly, the industry with the second highest expected increase is Financial Services & Insurance, with a 22 point jump that would significantly shorten the gap between it and the ICT industry which, despite a small expected increase (10 points), would still be number one according to their expectations.

Finally, it is also relevant to look into how Portuguese firms performed in each of the different blocks of the survey. For this analysis, the Adjusted DAI Score was used to control for the differences in questions left unanswered, since some blocks had a higher average of blank answers.

The Adjusted DAI scores by block, as seen in figure 4, reflect that there aren't any major differences in the distribution of digital maturity.

Standing slightly above the rest, the block with the highest average score was Strategy Driven by Digital (48.7 Adjusted DAI points) which may suggest that a considerable amount of Portuguese business leaders are aware of the importance of digital transformation and of having a detailed plan regarding how to achieve it successfully. Furthermore, this block also takes the prize in the expectations for 2022 with an average adjusted score of 72.2.

On the other end, the lowest scoring block was Digitize the Core (44.8), in part due to the presence of the two lowest scoring dimensions out of the 36 total: Industry 4.0 (average of 38.2) and Supply Chain (40.6). Both these dimensions included a strong component of automation, and Internet of Things (IoT), therefore suggesting that there is still a long way until Portuguese companies see these practices as the new normal.

4.2 Natural Language Processing Results

From the surveys collected, 430 companies had viable results for the Digital Acceleration Index open-ended questions. The total sum of their replies, after being preprocessed and tokenized, resulted in 54.053 keywords and 5.342 unique tokens.

It is of note, however, that approximately 35.000 of these keywords came exclusively from survey respondents accompanied by the Nova project team. In other words, 10% of respondents were responsible for over 64% of tokens.

In regards to the open-ended question inquiring on what firms considered to be their biggest challenge in achieving digital maturity, 543 companies submitted a response, totalling 3.261 keywords and 952 unique tokens. Of the 44 Nova accompanied respondents, 36 submitted a viable answer accounting for 687 keywords and 377 unique tokens.

4.2.1 Wordclouds

The tokens obtained through the preprocessing stages can now be parsed through a program in order to create wordclouds. Taking into account the peculiarities found in the textual data collected in the survey, the wordclouds will be divided into subsets that allow for logical comparisons.

In figure 5, we can see the wordcloud obtained from answers to all of the 36 main dimensions across the 430 companies that left at least one reply. There was an effort to remove several low-value high-frequency words - words that appeared too frequent to provide any contextual value that were not part of the compiled list of stopwords. Some other words, such as ”empresa/company”, ”cliente/client” and ”processo/process”, while also being vague and in part prompted through in questions, were nevertheless deemed relevant. This is especially true when taking into account the smaller subset in figure 6.

Figure 5: Digital Acceleration Index Survey Keywords for all dimensions (n = 430)

The output of the wordcloud in figure 5 should give us a view of the trends or keywords associated with Digital Maturity for all companies within the Portuguese context, namely how the biggest keywords point inwards towards company processes and words signifying internal change.

An analysis of the word frequencies (Appendix 1) further supports this analysis. Curiously enough, when compared to the wordcloud from the survey results excluding Nova-surveyed companies (Figure 6), the frequency of smaller words within the visualization reduces drastically. When analyzing these smaller keywords, which in large part can be attributed to the minority of Nova-surveyed companies that contributed the large majority of data, it is clear that keywords associated with strategy (e.g. ”objectivo/objective”, ”visão/vision”, ”management/gestão” and even ”estratégia/strategy” itself) have a bigger role in the replies to the 36 dimensions.

The collection of words in this wordcloud also seems to point towards the "Changing Ways of Working" block of the Digital Acceleration Index survey (see Figure 1), namely in regards to people, skills and processes.

In figure 6, we can see a similar wordcloud built from a smaller subset excluding the Nova-surveyed companies. It should be noted, once again, that this wordcloud featuring approximately 90% of companies actually represents a minority number of keywords and frequencies at just 40% of the total amount.

Figure 6: (Excl. Nova surveys) Digital Acceleration Index Survey Keywords for all dimensions (n = 386)

From this wordcloud, we can observe how nearly all major keywords are still featured prominently. The wordcloud is sparser in the number of words featured, despite having the same cap of 90 words.

A review of the word frequencies (Appendix 2) reveals that bigger words, despite maintaining their prominence, appeared twice or thrice times less. Additionally, medium-sized keywords also feature more prominently than in the figure 5 wordcloud. This can in part be attributed to the high number of companies in this dataset versus the lower number of keywords collected.

A further argument can be made that Nova surveyed companies tend to share similar attributes (Digital Maturity score, geographical location, size) and are in bigger contact with emerging digitalization trends, therefore offering more diversity and depth in their replies.

This effect is possibly compounded by data collection issues. In this subset of the data, company replies were shorter and sparsely populated across all dimensions.

The next two wordclouds were assembled from the responses to an additional Portuguese-context specific question regarding digital challenges faced more prominently by companies . Instead of putting together all the answers from the 36 main dimensions together into a bag of words, this question explores challenges in digitalization more specifically.

It should be also taken into account that most companies, including ones that volunteered no replies in any of the previous 36 dimensions, chose to reply to this one. The subsets of data chosen will analyze this time analyze the challenges faced by the Nova-surveyed subset of companies (n=37) when compared to the entirety of the data (527).

Figure 7: Respondent's Biggest Challenges in Digitalization (n = 527)

In figure 7, we can see the replies left by all respondents. Several of the keywords featured before through figures 5 and 6 appear once again. Keywords such as ”processo/process”, ”empresa/company”, ”transformação/transformation” are common in all four of the wourdclouds, and could be interpreted as companies facing internal challenges in regards to their digital transformation. There seems to be a bigger emphasis on challenges relating to the people and culture within an organization.

An analysis of the keyword frequencies (Appendix 2) nevertheless shows less overlap between keywords. From one wordcloud to another, certain words appear less frequently and are ranked less prominently at the top. The following keyword, in figure 8, should provide a stronger contrast in the more prominently featured keywords.

Figure 8: Nova Companies’ Subset on Biggest Challenges in Digitalization (n = 37)

Unlike before, the Nova-surveyed companies this time represent a minority of the datapoints token-wise (although they are nevertheless disproportionately bigger than they should be). This difference should be immediately noticeable when compared side-by-side. In figure 8, we can see a bigger prominence of medium-sized keywords, a fact that may be attributed to the lower number of companies (and a corresponding lower number of companies referencing the same 3 or 4 generic keywords).

Despite there being a higher level of entropy in the resulting visualization, smaller keywords do not feature as prominently when compared to these medium-sized keywords. They also seem to reveal that future trends in digitalization have to do with people and cultural aspects of the organization, with a lot of words focusing on these organizational aspects (e.g. ”transformação/transformation”, ”organizacional/organizational”, ”mudança/change”, ”talento/talent”, ”recursos humanos/human resources”, etc.)

Both of these wordclouds seem to indicate that companies all across Portugal will be facing challenges regarding organizational change, more so as they try to achieve greater levels in digital maturity. The challenges identified could also be related to other ”enablers” of the Digital Acceleration Index survey.

4.2.2 Company Network Graphs

Several adjacency matrices can be computed via matrix transposition and multiplication. All of them reveal, for the most part, the same result; an atomized and chaotic view of the nodes (which may be programmed to represent companies, dimensions and even individual keyword tokens themselves) with no underlying structure.

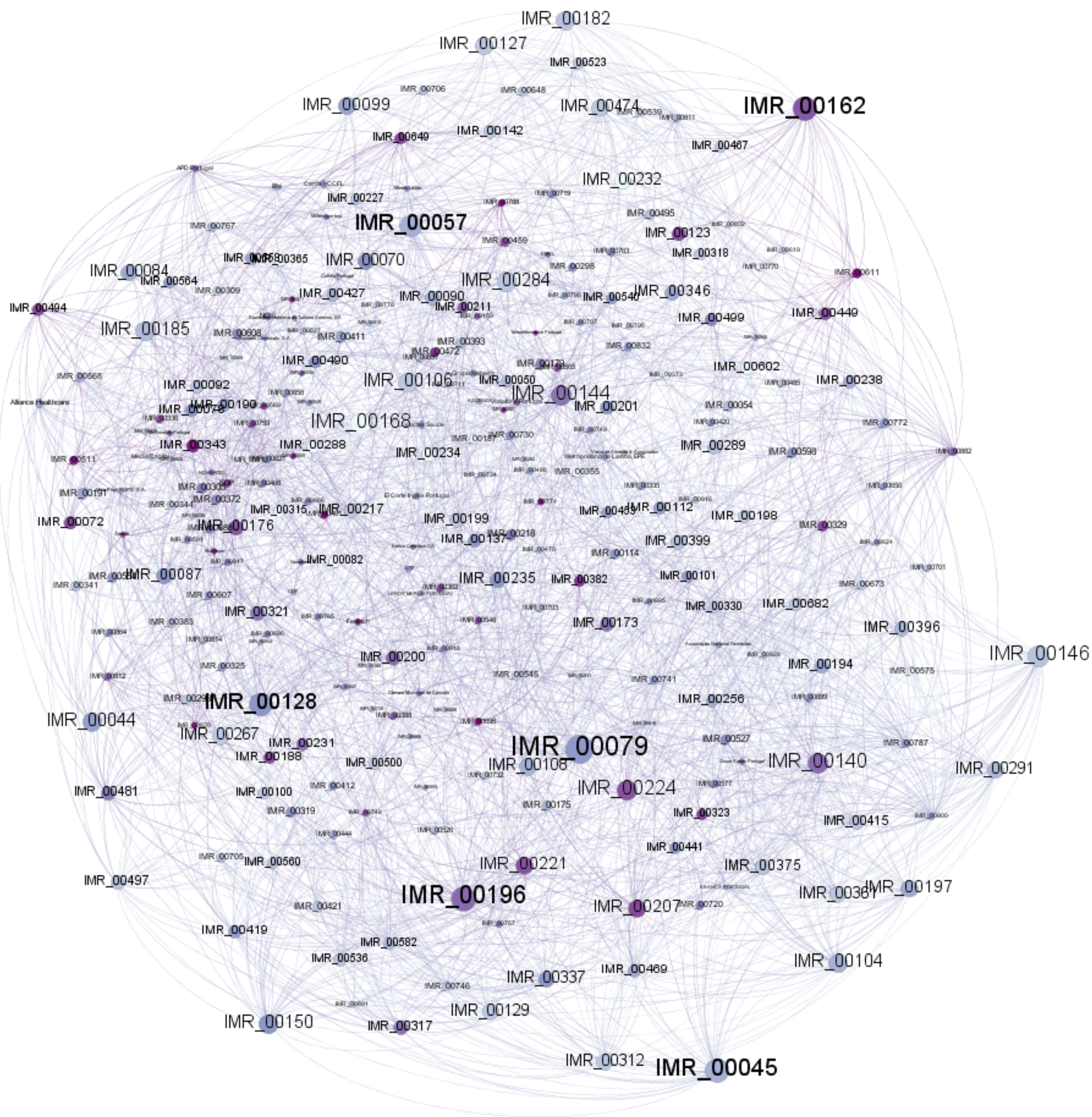

Figure 9: Company Networks built from vectorizing survey replies to biggest challenges in Digitalization (n = 247)

The above network was computed through company adjacency matrices. In other words, after creating a Term-Frequency Inverse Document Frequency (TFIDF) matrix, a company adjacency matrix was created through the process above described. The stronger an edge (as represented by its thickness) connecting two nodes, the stronger the similarity between tokens within a reply. As such, it can be surmised that two companies connected together through a strong edge might face similar circumstances between one another.

Furthermore, the node size represents the sum of the TF-IDF values for each company, meaning that the bigger the size of a node, the higher the likelihood that it either connects to more companies or that the contents within that company's reply are more relevant in the scheme of the entire survey replies. The color of the nodes represent a partitioning of the Digital Acceleration Index score, with darker shades representing a higher DAI score.

This network had its number of edges filtered by removing smaller edgeweights from an initial number of 10.216 just down to 2032, or just about 20% of all nodes. The edge weights are normalized and go from a scale of 0 to 1, with 1 being a perfect reply match in regards to the tokens utilized. From filtering the edges down to 20%, we preserved just the edges above 0.13 in weight. Regarding the number of companies in the network (274 versus a backdrop of 527 companies that submitted a reply), an initial rendition of this network did reveal a few clusters of companies. However, closer study of those clusters revealed outliers - companies that simply replied ”No answer” or that provided a single word reply. Upon removing those outlier responses, achieved by setting a threshold for reply length (e.g. only consider companies with at least 4 tokens after preprocessing in their replies) we arrived at the number of 274.

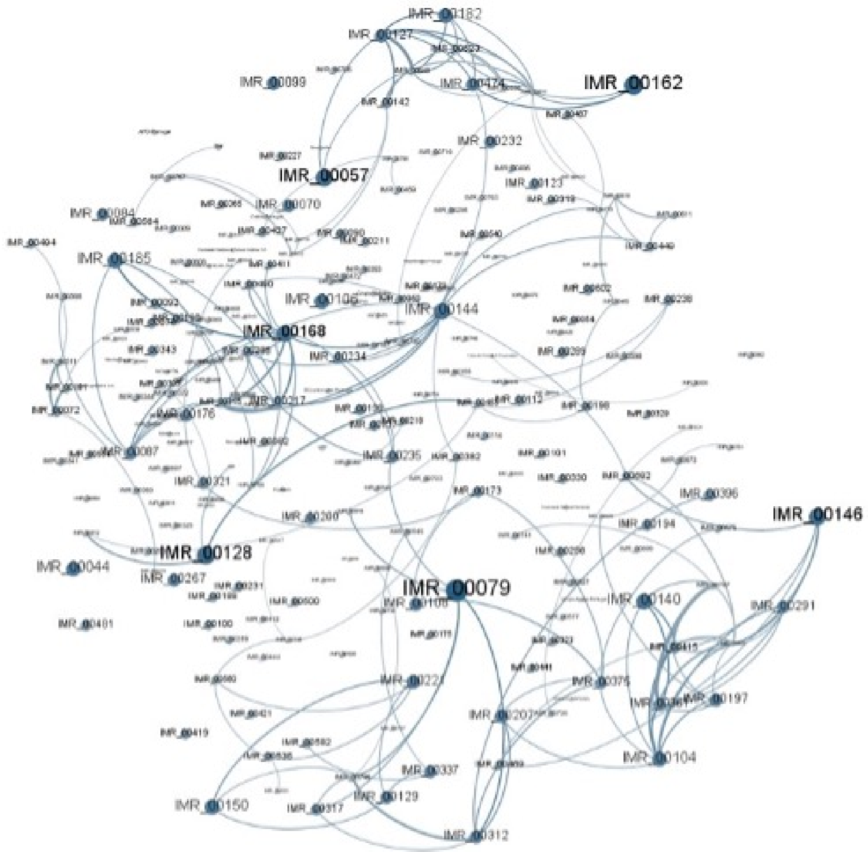

Figure 10: Company Networks built from vectorizing survey replies to biggest challenges in Digitalization preserving top 1.61% of edges ($n = 247$)

The above figure represents a further ”condensed” version of the prior network, this time preserving just the top 1.61% of edges.

A review of the individual company replies together with their direct neighbours with strong edgeweight reveals coherence; companies directly connected through a strong edge have similar replies or similar use in keywords. Although the network failed to reveal an underlying structure, even when choosing different node partitioning methods such as firm size, geographic location, sector, etc., the network provides an intuitive method for finding similar replies between companies, and could provide the basis for future consultancy work when analyzing such a large amount of textual data.

5 Discussion

This project aimed, firstly, to provide a starting analysis to the results of the digital maturity survey Digital Acceleration Index, developed by Boston Consulting Group and the first of its kind in Portugal. Secondly, it aimed to assess the relationship between the level of Digital Maturity and Wages within the Portuguese economy. Lastly, it looked to analyse trends digital maturity based on survey responses through a qualitative framework from text-based processing and data mining as an attempt at creating a machine based methodology to sift through big chunks of text. Due to the (national) scale of the project, the number of text-based questions, the open-ended nature of the questions and the breadth of the survey in regards to its dimensions and measurement, the insights gleamed from text-based replies were bound to be an invaluable, albeit under utilised asset.

The final sample of 797 firms in the DAI survey provided an average score of 40.79, slightly increased to 46.43 when adjusted for answers left in blank. This placed the Portuguese economy in a lower-intermediate level, classified as ”Digital Literate” by BCG's categorization. This result is in line with the findings from the European Comission (2020) which placed Portugal in a slightly below average position when compared to the European Union, mostly due to lagging behind in Human Capital digital skills. Responding companies were also asked for their expectations for improvement in the following two years and revealed a very optimistic outlook for such a short period, estimating an average increase of 16.5 DAI points, for a 2022 average expected score of 57.3. An analysis by firm size revealed that Medium and Big firms seem to have an advantage over Smaller ones. Separately, a sector analysis showcased a dominance of the ICT industry, while the Education sector proved to be the most confident in upgrading its digital maturity in the next few years. Lastly, a deeper dive into the results of different digital maturity topics present in the survey (here called blocks) did not show significant differences among them.

In regards to the model, the DAI stands out from the rest of practitioner models based on two traits. Firstly, due to its end to end range which encompasses the dimensions that are usually underrepresented in DMMs such as Business Model, Digital Ecosystems, Leadership, among others (Teichert, 2019). Secondly, the DAI differs from the majority due to its cyclical nature, given that it was designed to be repeated every 2 years and track the progress across each country where it is implemented. Moreover, one can argue that this is the first major practitioner DMM ever deployed in Portugal.

The Natural Language Processing methods, although falling somewhat short when faced with their initial scope, nevertheless utilized a throve of data that would otherwise would have gone to waste. A vast majority of the time spent by respondents in the survey is spent on these open-ended questions - a task that well surpasses the 4-hour mark. Furthermore, and in accordance with Boston Consulting Group's very own theoretical background, these survey replies should contain important information regarding present and future trends in Digital Maturity - trends that should go on to influence firms and markets in years to come, and that future editions of the should DAI@Scale Portugal should seek to prove or disprove.

The main contribution of the machine-based text mining methods used, from the preprocessing stages until the creation of meaningful and complex visualizations, is to provide a set of qualitative and quantitative tools for text analysis that would otherwise be impossible to process when taking into account the finite resources at researcher's disposal.

As for the keywords that accompany digital maturity trends in Portugal, the wordcloud and accompanying data should serve as an interesting benchmark or point of reference. Together with the network analysis, an in-depth look into the microtrends behind each of the dimensions or the macrotrends within each block of BCG's digital holistic framework could serve a standalone body of work for a deeper understanding of these trends in the Portuguese context across firms.

Finally, this project resulted in am important curated dataset, among the first of its kind, that will provide an important basis for future research done in the area of Digital Maturity.

This project cannot be viewed without taking some of its limitations into account.

Regarding the data collection, since it was not perfectly identic for all firms, there are a few biases that might come into play. Firstly, even though BCG and Nova School of Business and Economics catered to a small group of 44 responding companies by helping them fill out the survey correctly and ensuring that the right person or team in each firm was assigned the responsibility to answer, the remaining 1000 respondents were handled by the Instituto de Marketing Research and it is unlikely that such a big subset of firms had such a close walkthrough, if any at all. Naturally, this resulted in a higher rate of questions left unanswered, which not only harmed the quality of the data but also raises the question in certain cases as to whether the respondent was qualified enough to answer on behalf of his/her firm. Secondly, since it is a self assessment survey, there are psychological variants related to the employee appointed to respond, since some might be more optimistic/pessimistic in their expectations for the future and, in the case of high ranking managers, some might be averse to giving their own firm a bad rating or, in the case of bigger firms, might simply have a wrong perception of what really happens in the lower levels (Buvat et al., 2017).

The results from the Natural Language Processing also reveal several limitations in the survey. For starters, the disparity in data quality between the surveys collected by the Institute of Marketing Research and Nova's team left a very lopsided dataset, with a minority of companies (10%) accounting for the majority of text-based data produced (65%). Furthermore, the firms interviewed by Nova were primarily big in size, as well as major national players. As it stands, it is hard to draw any conclusions or establish any causal links between these companies and the rest of the economy in regards to digital trends.

Perhaps for many of the same reasons the network analysis did not yield any significant insights regarding its structure. Companies were equally distant and displayed similar centrality from one another. Neither were companies organized or clustered by any of the remaining factors measured in the survey, namely, firm size, geographic location, Digital Acceleration Index score, sector provided, etc.

Taking into account the survey's efforts into covering and detailing as many facets of a firm as possible, the fact that a network structuring the relationship between the Digital Acceleration's Index also did not yield any insights beyond a framework to study replies is also a point of contention.

In future developments on this topic, especially if further iterations of the DAI in Portugal do materialize in 2022 and 2024, it would be highly valuable to include the progression of digital maturity across time and follow up the study between Digital Maturity and human capital by Azevedo (2021). Additionally, since this Maturity Model was developed before the Covid-

19 pandemic's full impact, future repetitions of the survey could have specific fields about the agility and changes that the lockdown brought up. As for the text analysis, streamlining or reducing some of the questions might be in order in future iterations. Whilst the use of networks have been successfully used to study the relationship between keywords and theses, for example, the keywords from the survey respondents only reveals close connections between neighbours (as is the case of when two nodes are directly connected). This is useful as it allows a researcher to identify similar companies based on response similarity and find information, even if ultimately the direct neighbours come from a completely different sector and firm context.

Table 3: Appendix 1 - Frequency Tables 36-dim

	Freq. (IMR)	Freq. (Dataset)	Posição (Dataset)
digital	385	966	1
empresa	376	753	3
ainda	363	793	2
processo	282	741	4
falta	241	378	10
cliente	228	537	5
algum	213	402	9
investir	200	316	12
usar	188	245	25
digitais	178	451	8
pouco	164	210	31
inicial	164	176	47
recurso	159	13	316
fazer	159	252	24
área	157	486	6

Table 4: Appendix 2 - Frequency Tables for Challenges

	Freq. (Nova)	Freq. (Dataset)	Posição (Dataset)
digital	16	62	2
transformação	15	18	29
processo	13	40	10
empresa	13	54	4
mudança	13	24	26
tecnologia	11	34	14
organização	11	16	34
cultura	9	10	61
digitalizar	9	54	5
todo	9	31	17
negócio	8	28	19
projeto	8	13	40

References

Aslanova and Kulichkina (2020). Digital maturity: Definition and model.

Azevedo, J. (2021). Digital acceleration index: The relationship between firms' digital maturityand human capital.

Baig, A., B. Hall, P. Jenkins, E. Lamarre, and B. McCarthy (2020). The covid-19 recovery will be digital: A plan for the first 90 days.

Barata, J. and P. Cunha (2017). Climbing the maturity ladder in industry 4.0: A framework for diagnosis and action that combines national and sectorial strategies.

Bartel, A., C. Ichniowski, and K. Shaw (2007). How does information technology affect productivity? plant-level comparisons of product innovation, process improvement, and worker skills. *The Quarterly Journal of Economics 122*(4), 1721–1758.

Bloom, N. and N. Pierri (2018). Cloud computing is helping smaller, newer firms compete.

Brennen, J. S. and D. Kreiss (2016). *Digitalization*, pp. 1–11.

Bughin, J., L. LaBerge, and A. Mellbye (2017). The case for digital reinvention.

Buvat, J., B. Solis, C. Crummenerl, C. Aboud, K. Kar, H. E. Aouf, and A. Sengupta (2017). The digital culture challenge: Closing the employee-leadership gap.

Büyüközkan, G. and F. Göçer (2018). Digital supply chain: Literature review and a proposed framework for future research.

Comission, E. (2020). Digital economy and society index (desi) 2020.

Dalenogare, L., G. Benitez, N. Ayala, and A. Frank (2018). The expected contribution of industry 4.0 technologies for industrial performance.

Depaolo, C. and K. Wilkinson (2014, 05). Get your head into the clouds: Using word clouds for analyzing qualitative assessment data. *TechTrends 58*, 38–44.

Felch, V., B. Asdecker, and E. Sucky (2019). Maturity models in the age of industry 4.0 – do the available models correspond to the needs of business practice?

Fletcher, G. and M. Griffiths (2020). Digital transformation during a lockdown.

Frey, C. B. and M. A. Osborne (2013). The future of employment: How susceptible are jobs to computerisation?

Friedrich, R., M. L. Merle, F. Gröne, and A. Koster (2011). Measuring industry digitization: Leaders and laggards in the digital economy.

Gal, P., G. Nicoletti, T. Renault, S. Sorbe, and C. Timiliotis (2019). Digitalisation and productivity: In search of the holy grail – firm-level empirical evidence from eu countries.

Grebe, M., M. Rüßmann, M. Leyh, , and M. R. Franke (2018). Digital maturity is paying off.

Kannan, P. and A. Li (2016). Digital marketing: A framework, review and research agenda.

Laskina, L., L. Silakova, and O. Skidan (2020). The inter-relation between labor productivity and salary in the age of digital transformation: from economic analysis to the anthropological aspect.

Liu, J.-T., M.-W. Tsou, and J. K. Hammitt (2001). The impact of advanced technology adoption on wage structures: Evidence from taiwan manufacturing firms. *Economica 68*(271), 359–378.

Remane, G., A. Hanelt, R. C. Nickerson, , and L. M. Kolbe (2017). Discovering digital business models in traditional industries.

Summa, L. (2016). Digitale führungsintelligenz: ”adapt to win”.

Teichert, R. (2019). Digital transformation maturity: A systematic review of literature.

Tierney, P. (2012). A qualitative analysis framework using natural language processing and graph theory. *13*(5), 173–189.

Wamba, S., S. Akter, A. Edwards, and G. Chopin (2014). How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study.