Measuring Founding Strategy*

Jorge Guzman
Columbia University †

Aishen Li
Tsinghua University ‡

Abstract

We introduce a novel approach to measure the founding strategic differentiation of startups and its relationship to follow-on performance. We use natural language processing and historical websites to estimate the similarity between the founding website of an individual startup, the historical website of public firms at the startup's founding year, and the founding website of other startups founded in the same year. We propose that distance in the value proposition stated in these websites represents differentiation in the market. Startup differentiation is estimated as the average text-based distance from the five closest incumbents (public firms). We implement this approach using a large sample of startups from Crunchbase. Our measure predicts a meaningful increase in early stage financing and equity outcomes, unconditionally and controlling for cohort and industry fixed effects. The positive benefits of equity outcomes only evidence themselves after year 6 of age, suggesting more differentiated firms may take longer to prove themselves. Using out-of-sample tests, we also demonstrate that our measure is economically important, predicting 30% of the total variation in the receipt of early stage financing and 20% of variation in equity outcomes. Public datasets of our differentiation score and scraped website data are provided, together with open source code to replicate our approach in other settings.

Keywords: *Entrepreneurship , Strategy , Strategic Positioning , Machine Learning , Text Analysis*

*We thank Mabel Abraham, Eric Abrahamson, Natarajan Balasubramanian, Joel Brockner, Raj Choudhury, Bo Cowgill, Enrico Forti, Kathy Harrigan, Daniel Keum, Bruce Kogut, Kristina McElheran, Stephan Meier, Patryk Perkowski, Robert Seamans, Scott Stern, and Zhiyuan Yu for helpful comments, as well as participants in the Columbia Data Science Day, the NYU Workshop on Machine Learning and Strategy, and Strategy Science.

†Jorge Guzman: Uris Hall, 711. Columbia University. Broadway and 116th. New York, NY. 10027. jag2367@columbia.edu.

‡Aishen Li: Tsinghua University. las21@mails.tsinghua.edu.cn.

1 Introduction

Successful entrepreneurial strategies depend on both early stage firm positioning and experimentation under uncertainty (Porter 1996, Koning et al. 2019, Gans et al. 2019, Gambardella et al. 2018). At the core of positioning is a startup's strategic differentiation: whether it can stake out a unique value proposition in the consumer market. However, the role and relative importance of this founding differentiation, and positioning more generally, has been a matter of debate. While recent frameworks of entrepreneurial strategy posit an important role for early positioning (Gans et al. 2019, Siggelkow 2001), others, including leading practitioners, consider it hardly consequential (McGrath & MacMillan 2000, Reis 2011). The need for empirically assessing the strategic differentiation of startups and the relationship of this differentiation to performance outcomes appears evident. Yet, there is a key measurement challenge to do so: quantifying how unique startups are at their early stages requires observing all startups early in their lifecycle and developing a systematic way to understand their differences to the incumbent industry structure *at founding*.[1] How could one systematically evaluate founding differentiation, and positioning in general, for startups? What is the association of such measure with follow-on performance? More broadly, how can we measure a startup's founding strategy?

This paper introduces a novel approach to measure the founding differentiation of startups and assess its relationship to follow-on performance. To do so, we apply natural language processing methods to text written by startups and incumbents around the time of a startup's founding, and create measures of the differentiation of each startup from the incumbent industry structure. Concretely, using the WaybackMachine (an online historical archive of internet web pages) we download a copy of the website at or close to founding for a large sample of startups from Crunchbase, and the website of all public companies during the startup's founding year. We then use the word-embeddings algorithm doc2vec (Le & Mikolov 2014) to create vectors of embeddings that represent the use of text in each website and estimate the cosine distance between these

[1] Prior measures of differentiation have focused on using the presence of startups across industry-specific product categories, such as the number of categories in which video game startups release games (Cennamo & Santalo 2013).

embeddings vectors.[2] Using work in industrial organization as a heuristic, we aggregate distance measures to only consider the distance from the five incumbents closest to the focal firm.[3] We call this new measure the startup's *Differentiation Score*. Conceptually, this measure represents the distance between the value proposition stated by a startup on its website at founding, and the value proposition stated by other companies in the market. After documenting our measure, we empirically show it predicts performance outcomes and statistically accounts for an important share of their variance. Together, our method and results allow us to measure founding strategy and its correlation to performance. These results are accompanied by public code and data to use our measures and expand on our approach in other settings.

Before delving into further details of our method and results, we illustrate the intuition of how firm statements can be used to measure differentiation by considering how a human analyst may assess the level of strategic differentiation amongst firms. An analyst could do so by studying company marketing statements, where companies tend to emphasize their differences and unique value propositions. Consider Southwest Airlines and Delta Airlines. Southwest's slogan is "Low fares. Nothing to hide. That's TransFarency!". This slogan emphasizes low cost and transparency, which will be particularly appealing to cost-sensitive customers tired of extra fees. Delta instead uses the slogan "World's Most Trusted Airline", which does not focus on low cost but instead on trust and global coverage. Trust and global coverage might not be as valuable for the cost-sensitive travelers to which Southwest caters, but will be for those travelers that seek to get anywhere reliably and on time, and are willing to pay extra to do so.[4]

In this simple comparison, a strategy analyst can quickly and intuitively identify the differences between these two statements, even for companies in the same industry. These differences in statements do not simply reflect differentiation in the product features. The product in this case is ostensibly similar (a flight), and the statements instead capture the value proposition,

[2]We used the regularized version of the embeddings vector and remove the mean of each embedding column before processing (Mu et al. 2017).

[3]This heuristic comes from Bresnahan & Reiss (1991) who show only the first five competitors shape profits and (Igami & Uetake 2019) who show only the first five competitors shape innovation incentives. Our paper also studies distance to the one closest incumbent, five closest startups in the cohort, and one closest startup.

[4]In fact, Southwest has a relatively low rate of on-time arrivals.

be it variety (in route coverage, for Delta) or cost-leadership (Southwest). Even for companies whose competitive advantage is not a unique product, but the ability to deliver a lower price, as in Southwest, these features are emphasized in the company's marketing to consumers. Now, if the strategy analyst is given a third company—such as Spirit Airlines, which has the slogan "Less Money, More Go."—they will also recognize a difference in the perceived 'distance' between this new statement and the prior two. Spirit Airlines appears closer to Southwest, since both focus on the importance of low cost, and will therefore impinge on the differentiation of Southwest more than of Delta.

Our approach expands this idea to all companies and a richer set of company statements. If an analyst had a consistent source of marketing materials—such as the website— of a large sample of companies in the United States, would they be able to map systematically the conceptual 'distance' between the value proposition of one company and all others? Wouldn't a measure of this distance monotonically reflect higher (or lower) strategic differentiation? Building on this idea, we use natural language processing as a scalable tool to assess distance in the historical founding websites of startups and public firms and hence measure strategic differentiation at founding.

After developing our measure of strategic differentiation, we being validating it qualitatively by performing an in-depth look at the ranking of all startups in the Crunchbase category of Consumer Electronics. The overall pattern suggests our method maps to the construct of differentiation well. For example, in consumer electronics, companies that rank at the top of our differentiation measure include Zoi, initially developed as a wearables company that would use analytics to create a personalized running coach, and Healthy Stove, the world's first fully interactive oven. In contrast, the bottom includes more traditional companies such as Henge Docks, a manufacturer of docking stations for Apple products. When we look deeper into which are the closest public companies matched by our algorithm to these startups we can also easily observe how much more differentiated the top ones are. Zoi does match to public companies that are related to the overall area of health, fitness, and devices, such as the retailer Brookstone, which sells personal health devices, the sports retailer Dicks Sporting Goods, and Biozoom a medical devices

company. However, these matches are not close to the specific type of product Zoi sells. Henge Docks, on the other hand, has matched public companies that are also likely to sell products that can substitute away from Henge Docks's docking station, including electronics retailer BestBuy, and computer manufacturers AMX and Gateway. We then repeat this exercise for a very different category: Food and Beverage. We observe a similar pattern. The top company in this category is Coda Signature, a Denver-based startup developing cannabis-infused premium chocolates and truffles. This was a new market category that was not covered by any public company at founding. On the other hand, the bottom companies are much more traditional, such as the Olomomo Nut Company, selling nut products and snacks. We conclude from these validations that our measure captures the type of differences we would intuitively expect in a measure of founding differentiation and hence is a good representation of the underlying construct.

We next proceed to study how our measure predicts startup performance outcomes. While these estimates are not causal, they represent useful comparative statics of the relationship between founding strategy and eventual firm success and are both validations of our measure and new facts on the descriptive association between founding strategy and performance.[5] We focus on three key results.

First, we consider how founding strategy predicts early stage financing. We focus on regressions with founding year and industry fixed effects, where industries represent text-based industries replicating the approach of Hoberg & Phillips (2016). Compared to firms in the 10th percentile of our measure, firms in the 90th percentile are 10% more likely to raise early stage financing and raise 117% more total early stage financing.[6] In contrast—and consistent with the idea that we are capturing positioning vis-á-vis the consumer market and not the financing market—differentiation from other startups has much lower magnitudes and is negative when

[5]Specifically, while our analyses show that firms that report a higher level of differentiation early on are (statistically) more likely to succeed afterward, they do not imply making a more distinct website directly would change the likelihood of success. Rather, this success is possibly driven by other variables, such as founder human capital and intellectual property, that drive both better strategy formulation and eventual outcomes. Studying the causal relationship of changes in strategic differentiation to firm performance is left for future work.

[6]We define early stage financing as the sum of seed and pre-seed capital, grants, crowdfunding, and angel financing.

both types of differentiation measures are included in the same regression. Founding differentiation, particularly from the existing market structure, predicts the receipt of short-term financing.

Second, we use a similar specification to consider how founding strategy predicts long-term equity outcomes such as IPO or acquisition. The relationship here is more nuanced. Differentiation is positive over the startup lifecycle, but there is a significant age dependency. Indeed, consistent with the literature documenting that more unique startups will struggle to achieve legitimacy initially, but may ultimately perform better (Deephouse 1999, Marx et al. 2014), there is an initial negative relationship between our measure and the cumulative probability of equity outcomes that inverts after age 6, eventually predicting substantially higher outcomes for firms that have higher differentiation. Compared to firms in the 10th percentile, firms in the 90th percentile are 19% less likely to achieve an exit event in their year of birth (relative to the mean), but this changes to a positive cumulative difference of 8% by year seven, and 28% by year ten of age. This inverted pattern holds for both IPOs and acquisitions, with similar relative magnitudes, and it is even more striking for high-value acquisitions (i.e., over $ 100 million). Founding differentiation is a relevant predictor of equity performance.

Finally, third, we move beyond regressions to instead study the extent to which founding differentiation is economically relevant. Using out-of-sample ROC scores to assess the variance explained, we show that a fully interacted model of four measures of founding differentiation[7] predicts 30% of the variation in the receipt of early stage financing and 20% of the variation in the receipt of equity growth outcomes. Since our measures are naturally incomplete and imperfect, this estimate is best interpreted as a lower bound on the importance of founding strategy in this sample. Furthermore, if one takes seriously the argument in Teece et al. (1997) that high technology startups are the setting in which founding strategies should matter the least, this would also suggest that the extent to which differentiation predicts performance for firms in general is substantial. Founding strategy not only relates to performance but plays a meaningful role in statistically explaining variation in these outcomes.

[7] We use the differentiation from the five closest public firms, the single closest public firm, the five closest startups of the same cohort, and the single closest startup of the same cohort.

Through these results, this paper contributes to two distinct areas of the strategy literature. The most important contribution is methodological. This paper provides a formal way to measure founding strategic differentiation, building from the tenets of strategy. The new measure we propose is different from prior attempts at formalizing strategy in that we focus specifically on how a firm is occupying a distinct value proposition from its competitors, as reflected in its marketing. While prior work has instead sought to define the nature of what strategy is (Van den Steen 2016, Porter 1996), there are many applications where measurement itself is of fundamental value. We hope that our results provide useful guidance to researchers seeking to measure differences in the level of strategic positioning and startup founding differentiation. We suspect follow-on work will improve upon our approach. To support this effort, we have released through our Github repository all our code as open source and the vast majority of our data, including differentiation scores for startups, the raw website text used, and the trained word-embedding models that allow assessing the similarity of new firms to ours.[8]

Our second contribution is to the entrepreneurial strategy literature. Our results provide novel evidence on the importance of founding differentiation, and founding positioning, to performance outcomes. While prior work has theorized that high technology companies such as those in Crunchbase may gain little from founding positioning, and instead achieve performance through either dynamic capabilities or experimentation (Eisenhardt & Martin 2000, Teece et al. 1997, McGrath & MacMillan 2000, Reis 2011), we show that this perspective may be too extreme. Rather, the essence of startup strategy may require recognizing both positioning and experimentation (Siggelkow 2001, Gans et al. 2019), and how they work together to develop a competitive advantage in ways that is sustainable, but also quickly adaptable to rapidly changing contexts. Understanding better the interplay of static positioning and dynamic capabilities and experimentation is an important area for future work.

Perhaps the paper closest to ours is the influential work of Hoberg & Phillips (2016) (hereafter HP16). HP16 use the text in the business description section of the annual reports of public

[8]These are available at https://bit.ly/MeasuringFoundingStrategy.

firms to develop a new approach to understand industries and their dynamics. To review, HP16 use the cosine similarity between word vectors weighted by the term frequency-inverse document frequency algorithm (tf-idf) to estimate a text-based distance between firm statements. Then, they implement clustering algorithms and propose each cluster represents a different text-based industry, ultimately showing that these industry definitions describe industry dynamics significantly better than SIC codes. Relative to this work, our paper offers several novel contributions.

First, conceptually, our paper's focus is on strategy rather than industry. This means that while HP16 focus on how companies agglomerate into groups of related firms, strategy focuses on what makes a company distinct from all potential competitors. Understanding those business-specific elements that drive firm performance beyond industry is at the core of strategy research (Rumelt 1991, McGahan & Porter 1997, Ruefli & Wiggins 2003). The precise construct we measure, strategic differentiation, is not measured in any of the papers by HP16, nor is our analysis of how these measures predict startup financing and equity outcomes, or our estimates of the economic importance of founding strategy for technology-based startups.

Second, methodologically, all our analyses include fixed effects for the text-based industries defined by HP16—created by replicating their methodology within our data—and our regressions cluster our standard errors by these industries. This means that all our effects are effectively estimated conditional on the HP16 industries.

Third, also methodologically, our machine learning model implements a more sophisticated natural language processing method than HP16. Our algorithm uses word-embeddings, while HP16 used cosine similarity of relative frequency (tf-idf). The key difference between the two is that the word-embeddings approach also incorporates the context in which words are used when creating vectors to describe text documents. Concretely,[9] while tf-idf counts the presence of each word (or its stem) weighted by how infrequent the word is, word-embeddings runs a neural network that tries to predict a word based on the words that surround it to develop a vector of

[9]In our implementation, we use the doc2vec expansion of word2vec developed by Le & Mikolov (2014) which allows developing embedding vectors for whole documents. However, we limit this explanation to the simpler word2vec model for ease of exposition.

weights representing each word (in our case, for each word, we use a window of the seven prior and seven follow-on words). This use of context makes word-embeddings algorithms give two synonyms similar weights if they are surrounded by similar words, and the same word get very different weights if it is used in a different context. For example, the words 'social media' would receive very distinct weights when they are new, since even though both 'social' and 'media' are common words they had not been used together in this context, and are likely surrounded by other words that didn't tend to be neighboring them before. However, the words 'Facebook' and 'Twitter' will get an embeddings vector similar to each other and to other social media words, since they are often used in the same context. Introducing the role of context turns out to be important in our analyses. In a parallel to the idea in strategy that successful positioning requires not only using novel elements but also combining them in unique ways, we find that the tf-idf model has a very negligible role in predicting equity outcomes, accounting for only 3% of all variation, while the word-embeddings model, in contrast, accounts for 20%. We conclude that, in our setting, our measure is more meaningful in an economic sense.

Finally, fourth, from a data perspective, our paper is also the first to focus on startups and to do so using their founding websites. Because websites are usually marketing tools,[10] while the 10K statements used by HP16 are investor-oriented, our approach is better able to capture market differentiation in itself, while HP16 captured the top management team's perspective on the competitive landscape. Understanding and unpacking further the differences in what is conceptually captured in different firm statements, and how they represent different elements of firm strategy and disclosure, is a rich avenue for future work.[11]

The rest of the paper proceeds as follows. Section 2 presents our formalized methodological approach. Section 3 reviews our data. Section 4 presents our results. Finally, Section 5 concludes.

[10]While websites can potentially do other things, reaching customers to explain the product is one of their core and most important functions. For example, in a column titled 'Why Every Business Needs a Website' in Forbes Magazine's small business section, the listed reasons are i) credibility, ii) brand, iii) leads, iv) organic traffic, v) customer service, vi) announcements and vii) digital marketing. Except for customer service, all others are core marketing functions (Forbes 2020).

[11]In related work, Boulland et al. (2019) show that the size of a website of public firms and the specific url headers are related to the material disclosures these firms make to investors.

2 Measuring founding strategy: a text-based approach

Our approach is anchored around the idea that the relationship between text written by firms can be informative about the underlying market structure (Abrahamson & Hambrick 1997, Hoberg & Phillips 2016). This section overviews how we use firm websites to assess similarity in the value propositions of firms, how we translate this similarity to a measure of distance, and how we aggregate this distance into a measure of strategic differentiation, allowing us to score startup differentiation at founding.

Measuring market differentiation through firm statements

Our approach to measuring strategic differentiation builds on four insights. First, while it is virtually impossible to observe the value a consumer sees in a product, it is possible to observe what the firm believes its value proposition to be. Firms constantly state their value proposition in their marketing statements to explain to consumers (or to some representative set of them) why their product or service should be purchased. Second, the similarity in these firm marketing statements is a good indicator of the substitutability between the value proposition of their offerings, thus allowing the assessment of how differentiated is a new startup from incumbent firms. Third, measuring distance between company statements is not merely a theoretical idea: there are standard text-analysis algorithms that allow us to quantify the relatedness of those statements to effectively create a measure of similarity in the stated value proposition of firms in the market. Distance is then simply the inverse of similarity. Fourth, observing at least some of these statements at or close to founding is possible through the use of archival websites. Since websites represent a de-facto marketing channel for virtually all firms founded after a certain date, the distance between founding websites can be used to measure founding positioning.

Building on these insights, we define market relatedness as a measure representing the similarity between two firm statements. Given a startup and an incumbent statements s_i and s_j (one each) explaining their main value proposition, there exists some function h defined between 0 and 1 that can measure a pair-wise similarity between these two statements as

$$\sigma_{ij} = h(s_i, s_j), \sigma_{ij} \in [0, 1]$$

Companies with a value of similarity equal to 1 have completely equivalent statements, while companies with a similarity value of 0 have no relationship to each other. Companies with partial similarity are in between.

Implementing word and paragraph embeddings

To define the similarity function h, we focus on a specific natural language processing (NLP) algorithm called word-embeddings (word2vec) (Mikolov et al. 2013). Compared to traditional bag-of-words approaches such as the term frequency-inverse document frequency algorithm (tf-idf) or topic modeling, the distinction of word-embeddings algorithms is incorporating the context in which words are used when characterizing them.[12] In essence, while in tf-idf a word is weighted only by how uncommon it is across documents, word2vec represents each word through a vector of N factors (embeddings), creating a factor-based description of the word. These factors are estimated through a neural network that predicts the probability a word is used based on other words that occur before or after it. Doing so means that when the same word is used in a very different way (such as Casper, the friendly ghost, and Casper, the company) it results in a very different vector, while if a word is synonymous to another one (such as Nectar, also a mattress company, versus Casper, the mattress company), a similar vector would be estimated even though they are spelled completely different. In contrast, in this example, tf-idf would have delivered the opposite conclusion: scoring the two versions of Casper as the same, and different from Nectar. In our implementation, we use the expanded version of word2vec, doc2vec (Le & Mikolov 2014), which allows us to take advantage of this technique to build document-level vectors. Finally, building from the insights of Mu et al. (2017) we include a post-processing step and subtract the sample mean of each embedding to itself, to make all of them mean zero. $h(s_i, s_j)$ is estimated as the

[12] Bag-of-words approaches also allow incorporating semantics by increasing the number of words in each n-gram, however, this tends to quickly lead to sparsity.

cosine similarity between the normalized version of the embeddings vector of any pair of firms i, j.[13]

From similarity to founding strategy

The next step is to aggregate the pair-wise similarity between all startups i and incumbents j into a firm-level measure of differentiation. We first define distance, δ_{ij}, by algebraically inverting σ_{ij}. Distance is a value between 0 and 1, where 0 indicates that two companies are the same and 1 means they are completely different.

$$\delta_{ij} = 1 - \sigma_{ij} \tag{1}$$

Next, we aggregate distance across all incumbents to get an empirical measure of the differentiation score at founding. The mean or median are not good ways to aggregate measures of competitive overlap because most companies are unrelated to each other. Empirical studies in industrial organization highlight how the dynamics of competition are influenced by a small number of competitors and how, as this number increases, the ability of firms to charge margins quickly decreases, approximating a fully competitive economy (in strategy parlance, they lose their competitive advantage). We follow a simple heuristic and use the classic finding of Bresnahan & Reiss (1991) showing that markets become competitive after the first three to five competitors.[14] While this heuristic is admittedly ad-hoc and imperfect, it allows a tractable approach that is applicable across many firms.

The differentiation score is:

$$\hat{S}_i = \frac{1}{5} \sum_{j \in J_i^5} \delta_{ij} \ , \ J_i^5 = \{5\, closest\, incumbents\} \tag{2}$$

[13]Our implementation uses the the gensim.models.doc2vec Python library with the following parameters: a vector size of 700, a word window of 7 (both before and after the focal word), ignore all words that occur less than 3 times, using a distributed memory algorithm (PV-DM), no corpus file, and 10 total iterations over the corpus (epochs). All code is available at https://bit.ly/MeasuringFoundingStrategy.

[14]In more recent work, Igami & Uetake (2019) also study the impact of competition on incentives to innovate and find that incentives to innovative drop quickly, stabilizing after five competitors.

We also estimate a startup's differentiation from the single closest incumbent, the five closest startups with the same founding year, and the single closest startup with the same founding year.

3 Data: Crunchbase, the Wayback Machine, and industry controls

We implement this approach on a comprehensive list of startups from Crunchbase, which we complement with their historical websites at the time of founding, and the annual websites of publicly listed firms in the United States. We also include the industry of each startup, estimated by replicating Hoberg & Phillips (2016) within our data. We describe each dataset below.

Crunchbase startup data. We obtained data on all companies available in Crunchbase founded between 2003 and 2019 that have a website. Crunchbase is a popular crowd-sourced data platform tracking a large number of technology-based startup companies. It is one of the main databases used in entrepreneurship and strategy research and performs particularly well in covering innovative firms that receive some form of institutional financing (Dalle et al. (2017) provide an overall assessment and examples of the use of Crunchbase in management and economics research). The data includes both active and deceased companies and we included all firms of any status in our analysis.[15]

For each company, we downloaded in April 2019 the company name, the founding date, the website address, the city and state of the main office, the date and amount of each financing round, whether the company achieved an equity event (IPO or an acquisition), the market valuation of the company at the exit event, the timing of the exit event, and the top-level Crunchbase category for this firm. Crunchbase categories are conceptually industry categorizations, but their focus is on characterizing firms across groups that better delineate startup industries than traditional SIC codes.

Website history data with the Wayback Machine. We used the Wayback Machine, an online

[15] As a matter of policy Crunchbase keeps all companies ever recorded, except for special circumstances (see Quora (2013) for further explanation).

platform offered by the Internet Archive (archive.org), to download the initial website of each startup around the time of founding. The Wayback Machine provides access to a digital library containing over 330 billion web-page snapshots occurring in history. These snapshots are taken at least a few times a year for all unique domain names on the internet. We developed a web-scraping technology, available in our Github repository, to automatically query the Wayback Machine for the earliest version of the webpage in the year after the year of founding in Crunchbase. We downloaded the homepage and the first level links in the webpage (up to ten URLs to limit the size of the download). We excluded all pages that returned empty, that included too little text, that were not in English, that were not of MIME types 'text/html' or 'text/plain', that reported an HTTP error such as a 403 or 303, or that appeared to be a boilerplate such as the default page of an Apache web server.

Incumbent information. To consider the existing market structure at founding we focus on publicly listed firms. Specifically, using the IPO and de-listing dates in Preqin, we downloaded the first available website each year for all companies publicly listed in NASDAQ and the New York Stock Exchange using the same download algorithm used for the startups. This allows us to observe the market proposition of all public companies as stated at the time of startup founding and thus assess adequately the startup's positioning in the market at this time.

Industry controls. Finally, we develop industry categorization by implementing the method of Hoberg & Phillips (2016) (hereafter, HP16) to develop clusters of related industries based on the company's own business descriptions within our data. To review, HP16 use the term frequency inverse document frequency (tf-idf) algorithm in the business description of 10K annual reports to develop vectors of weighted words, and then the cosine similarity between these vectors to estimate a scalar distance from one company to another. Then, they implement a k-means clustering algorithm and use the resulting cluster identifiers as the industry categorization. HP16 recommend using 300 clusters as the target number to mimic well the distribution of SIC industries.[16] We implement this method with 300 industries using our website text rather than the 10K

[16]This algorithm is the fixed-industry classification algorithm in Hoberg & Phillips (2016) (p. 1435). Hoberg & Phillips also implement a 'network' based measure which our approach does not allow us to implement.

business descriptions, to define industries within our data. The resulting variable *HP Industries*, represent 300 indicators for the clusters created through this method. The median number of startups in an industry is 36, and the average 50.

Estimates of similarity and founding strategy

We call our preferred measure *Differentiation Score (5 Closest Public Firms)*, defined as the average distance from the focal startup's website at founding to the five closest public firms. Figure 1 reports the full distribution of our measure. As is apparent, there are some outliers that have a significantly low differentiation score. To avoid having outliers drive our results, and instead focus on the core correlations of the data, we winsorize the distribution.[17] Table 2 reports summary statistics for four estimated differentiation scores, after windsorizing. Our preferred measure has a mean value of 0.64 and a standard deviation of 0.064. The difference between the 10th and 90th percentile is 0.17. We also include measures for the distance from the single closest public firm, the average distance from the five closest Crunchbase startups founded in the same year, and the distance from the closest Crunchbase startup founded in the same year. Table A1 reports the correlation between these four measures. Measures based on public firms and measures based on startups have very high correlations to each other, greater than 0.9, while the correlation between measures in these two groups is slightly lower, ranging from 0.59 to 0.74.

Summary statistics

Table 1 presents summary statistics of our data. There are 12406 startups in the data. The average founding website length for a startup is about eleven thousand characters, but this measure is significantly skewed. Moving to outcomes, 68% of firms in Crunchbase receive early stage financing, which we define as receiving any of angel financing, grant financing, crowdfunding, and seed and pre-seed rounds. The average early stage financing received is $909 thousand. 29% get series A financing, with an average investment of $2.1 million. 18% of firms achieve an equity growth

[17] Specifically, we exclude the top and bottom five percent. All results reported in this paper are robust (and often even higher and more statistically significant coefficients) when including these outliers, but the validations looking at the distribution of firms in the Consumer Electronics category suggested some firms at the bottom tail matched due to invalid text and error messages, rather than actual firm text.

outcome, of which IPO represents only 1.7% and acquisition is 16%. There are 1.5% firms which have a reported sales price of at least $100 million (about 10% of all acquisitions). We categorize these as high-value acquisitions.

4 Results

Qualitative validations of the estimated differentiation score

We begin by considering specific examples of how our measure describes companies within industries. To do so, we introduce in Appendix B the full list of all companies in our data for two Crunchbase categories, their differentiation score, and the startup's description as written in Crunchbase. We emphasize that both the Crunchbase description and categories are independent of the website text we use to develop our measures, and not used anywhere in our approach estimating similarity or the follow-on regressions.

Table B1 is all companies in Consumer Electronics. We focus on a few examples in the extremes. The top-ranked companies appear innovative. For example, Zoi (now Runteq), ranked at the 98th percentile of our measure, was a company developing wearables to do personalized running coaching through sophisticated data analytics, and Healthy Stove, also at the 98th percentile, described as the world's first fully interactive oven. These appear, by and large, innovative consumer electronics products. Indeed, when the online magazine SportTechie covered Zoi, it highlighted particularly their distinctly innovative product ("Zoi is the first option to provide a virtual coach") and its promise ("We look forward to seeing future generations of this concept").[18] Table B3 delves deeper into our measure by reporting the public companies that score closest to Zoi in our algorithm. These include the health and device retailer Brookstone, the sports retailer Dicks Sporting Goods, and the medical device company Biozoom. By and large, they appear adequately related to Zoi as they operate in the broad areas of devices and health. However, they are also meaningfully different, and hence Zoi scores high in our differentiation score.

At the bottom of the Consumer Electronics category we find a contrasting set of products

[18] https://www.sporttechie.com/wearable-technology-and-the-way-we-run/

that are more typical of existing offerings, and hence less differentiated from incumbent firms. These include, CorasWorks, a company developing websites using Microsoft Sharepoint technology, and Henge Docks, a company offering a docking station for the Apple Macbook. Apple itself does not offer a docking station for the Macbook, making Henge Docks's product a valuable one, but it is also in clear ways less distinct than Zoi. Indeed, when AppleInsider reviewed Henge Docks's product it did not emphasize its distinctiveness or innovativeness, calling it simply "an expensive and elegant way to work at your desk".[19] In Table B4, we see that closest public companies matched to Henge Docks are also companies that sell other docking stations and computer products including the electronics retailer BestBuy, and the computer manufacturers AMX and Gateway. Generally, these appear to be more close competitors to Henge Docks, than Zoi's matches. Hence, there is a lower differentiation estimate for Henge Docks to the ongoing market structure.

We repeat this exercise in Table B2 but focus on a different type of category, Food and Beverage. The top companies once again include well-differentiated new ideas such as Coda Signature, a creator of cannabis-infused chocolates and high-end truffles, and Ripe.io, with the slogan 'the blockchain of food'. Consistent with the idea that Coda Signature's value proposition is innovative, it was awarded the Excellence in Innovation Award in 2019 by the National Cannabis Industry Association.[20] Comparing it to public companies, Coda Signature is markedly distinct. While it matches to other public companies that may sell high-quality chocolates and health-oriented food, such as the supermarket Publix, the coffee chain Caribou Coffee, and the beverage company Pulse Beverage, the differences between these products and Coda Signature's is substantial. While cannabis is at the core of Coda Signature's value proposition, none of the public companies operate with any kind of mind-altering substance. As in Consumer Electronics, we once again observe at the bottom of the list firms that are more typical in this industry, including the Olomomo Nut Company, a traditional producer of nut products, and Saucey Sauce, a company

[19] https://appleinsider.com/articles/20/07/21/review-brydge-vertical-dock-is-an-expensive-and-elegant-way-to-work-at-your-desk

[20] https://codasignature.com/press-release/coda-signature-wins-2019-ncia-excellence-in-innovation-award/

creating new sauces and marinades.

Founding differentiation and early stage financing outcomes

We next assess the association between our measures and startup financing. Figure 2 presents binned scatterplots correlating our main measure, the differentiation from the five closest public firms, to early stage financing. Panel A includes only year of founding fixed effects. The relationship is positive and very precise. Within startup cohorts, startups with a higher differentiation score raise a higher amount of early stage financing. Panel B replicates what will become our preferred specification, introducing both founding year and HP industry fixed effects. The pattern is slightly less pronounced but still meaningful. This relationship holds even within groups of related competitors. Panel C considers the possibility that there is something about the way the websites are built that correlates to both our measure and outcomes by controlling for the length of website text. To do so, we split our variable of website text length into 20 bins and include them as additional fixed effects. Reassuringly, there is little change in the relationship. Finally, Panel D reports the extensive margin, whether a startup gets financing at all, with a similarly positive result. Appendix Figures A1 and A2 consider other differentiation scores, such as those from one closest public firm or the closest startups, and include outcomes that incorporate series A financing events as early stage. The results are very similar. These graphs suggest significant positive relationships between our measure, estimated close to founding, and whether and how much early stage financing startups get.

Regression Estimates. We report the relationship of our measure to financing more precisely in Tables 3 and 4. In Table 3, we study the extensive margin of financing by reporting an OLS regression with *Gets Early Stage Financing* as the dependent variable and the differentiation score as the independent variable. Standard errors are double clustered by HP industry and state to account for industry or location correlation in the error term. Column (1) shows a positive unconditional coefficient of 1.614. Column (2) shows that there are large cohort effects: the coefficient drops to 0.664 after including founding year fixed effects. Column (3) is our preferred specifi-

cation, which includes founding year and HP industry fixed effects. The coefficient is 0.402.[21] To put this result into perspective, this implies that relative to firms at the 10th percentile of our measure, firms at the 90th percentile are 6.8 percentage points more likely to raise early stage financing (about 10% of the mean). Finally, Column (4) is an additional robustness test that includes city fixed effects and state by year fixed effects to account for the possibility of geographic time-varying unobservables driving our effect. Our coefficient is very similar.

Next, in Table 4, we study the total amount of financing received, by using *Log(Early Stage+1)* as the dependent variable. The differences are more dramatic. Columns (1) and (2) show that there are similarly large cohort effects in our data. The coefficient for our preferred specification is column (3), using founding year and HP industry fixed effects. The estimate is 4.545. This implies that, on average, firms at the 90th percentile raise 117% more early stage financing than those at the 10th percentile[22] and that moving by one standard deviation in our measure predicts 33% higher early stage financing. Column (4) shows that this is robust to controlling for geography by including city and state by year fixed effects. We perform additional validations on columns (5) through (9). Columns (5), (6), and (7) disaggregate early stage financing into seed financing, grant financing, and angel financing. All coefficients are positive. Seed and grant financing are significant, while angel financing is slightly below the usual significance levels (p=0.13). Column (8) includes series A financing events together with early stage financing. The coefficient is positive with p-value of 0.11. Finally, column (9) reports a regression using series A financing only for firms that did not raise early stage financing. The coefficient, while positive, is noisy and far from significant. We conclude from the evidence in Tables 3 and 4 that the relationship of founding differentiation to early stage financing is positive, meaningful, and robust.

Other Differentiation Scores. Next, Table 5 considers the relationship of other differentiation scores to early stage financing. Column (1) repeats the preferred estimate of Table 4 for comparability, using the differentiation score estimated from the five closest public firms. Column (2)

[21] This drop of 40% when considering the explanatory power of industry on performance is lower than the classic estimate in Schmalensee (1985) on the role of industry to firm profitability, but higher than the follow-on estimates in Rumelt (1991) and McGahan & Porter (1997).

[22] The 10th-90th percentile range is 0.17; $e^{(.17*4.545)} - 1 = 1.17$.

instead uses the differentiation from the single closest firm. The coefficient is smaller at 3.203. Columns (3) and (4) focus on a different type of differentiation, differentiation from other startups in the same cohort. While differentiation from public firms is intended to capture the market positioning in the extant U.S. economy, differentiation from other startups may better reflect the uniqueness and positioning in the venture financing market or access to other startup resources. Interestingly, these coefficients are only a third in magnitude from our main effect. The differentiation from the five closest startups has a coefficient of 1.282, while the differentiation from the closest startup has a coefficient of 1.478. Columns (5), (6), and (7) introduce multiple measures at the same time to consider the correlation of one measure of differentiation conditional on others. In column (5), we note that when both differentiation from the five closest and single closest incumbent are included, only the former remains positive and significant. It seems including five incumbents rather than one gives more precision to our measure. More strikingly, when differentiation from both incumbent and startup firms is included simultaneously in columns (6) and (7), the incumbent measure remains positive and significant while the startup measure turns either negative or not significant. Together, we interpret these results as emphasizing that the positive relationship of our measure to early stage outcomes is driven by underlying differences from the existing market structure, as proxied by public firms, rather than differences from startups or its implications on competition in the financing market itself. Strategic differentiation in the consumer market predicts performance.

Founding differentiation and equity outcomes

We proceed to study how our differentiation score predicts long-term firm equity outcomes such as IPO or acquisition. Figure 3 reports binned scatterplots of our differentiation score with *Equity Growth*, a binary measure equal to 1 if the firm achieves IPO or acquisition, as the dependent variable, and *Differentiation Score (5 Closest Public Firms)* as the independent variable. Perhaps unintuitively, all relationships appear noisy and weak and have a negative slope in the fitted line.

Main Relationships. We study these relationships in more detail through regressions in Table 6. The coefficients show a different pattern than early stage financing. Column (1) reports a noisy

relationship of founding differentiation and equity outcomes within cohorts. This turns large and significant in column (2), using founding year and HP industry fixed effects. The coefficient is -0.112, and remains after including geography controls in column (3). On average, firms that score at the 90th percentile of our measure are 1.9 percentage points less likely to achieve a growth outcome than those at the 10th percentile (10% of the mean). Columns (4) and (5) separate IPOs and acquisitions. Column (4) shows a negative and meaningful coefficient for IPO, while column (5) reports a noisy but negative coefficient for acquisitions. Within our whole sample, the relationship from strategic differentiation to equity outcomes appears negative.

In Table 7 we perform additional analyses to provide more detail on this relationship across age and exit value. Column (1) drops all firms that are young in the data, and for whom, consequently, we cannot observe their whole lifecycle. To do so, we remove all firms born in 2012 or later. The coefficient is now positive, but noisy and close to zero. Columns (2) and (3) separate the exits based on timing. Column (2) is the exits early in the startup lifecycle, within the first six years, while column (3) is those after six years.[23] Once again, while the coefficient for the early exits is negative, the one for the late exits is not and has a positive, but noisy, estimate. These timing differences foreshadow an additional analysis is needed: the relationship of differentiation to outcomes may change over time, and hence we would need to consider the dynamics of firms across their whole lifecycle.

Columns (4) and (5) consider instead the valuation of the acquisition outcomes to study heterogeneity in how 'successful' an acquisition is. While Crunchbase only reports acquisition values for a fraction of acquisitions, we assume that the misreporting is not correlated with differences in founding differentiation. Column (4) shows a large positive effect for differentiation on the acquisition price, conditional on being acquired. Column (5) shows that when we consider only high-value acquisitions, instead of all acquisitions, the coefficient turns to zero instead of negative. Overall, there appears to be evidence that the relationship between differentiation and equity outcomes is also more positive when we consider 'larger' acquisitions.

[23] 58% of the startup exits in our sample occur within the first six years.

Dynamics. So far, we have limited ourselves to cross-sectional regressions, but this is unsatisfactory for equity outcomes given the regression results in Tables 6 and 7 showing the role of founding differentiation on outcomes may vary across firm age. Not simply an empirical regularity, the existing literature also suggests the dynamics of acquisitions and IPOs may vary depending on how differentiated a firm is at founding. One particularly natural hypothesis is that highly differentiated startups may take longer to develop as they face additional legitimacy costs in the market (Deephouse 1999). Indeed, Marx et al. (2014) show that firms that have disruptive technologies (which tend to be more differentiated products) develop dynamic commercialization strategies that require the firm to compete to eventually get to acquisition outcomes. In this case, for example, the less disruptive companies would be *more* likely to be acquired early on, but this pattern changes as the firm ages and disruptive companies achieve necessary legitimacy. For our analysis, this type of dynamics would create a negative bias, since many firms are only observed for a few years (e.g., the younger cohorts) and we do not get to observe the later years in the life cycle.

We study this possibility in Table 8. To do so, we estimate our preferred specification in a panel format and report the individual coefficients for founding differentiation score against the cumulative probability that a firm has observed an equity exit by each year of age. The pattern we see is dramatic and consistent with the dynamics of disruptive companies. Firms with a higher differentiation score are initially less likely to exit, particularly during their first year (year 0). For our main growth outcome, a firm scoring at the 90th percentile of our measure is 19% less likely (relative to the mean) to achieve equity growth in year 0 compared to one at the 10th percentile. Yet, this pattern reverses as the firm ages. The coefficient turns positive by age 6 and continues increasing thereafter. By age 7, firms with a higher founding differentiation score are more likely to have had an equity exit. The coefficient, with a value of 0.086, implies that moving from the 10th to the 90th percentile is associated with a 1.4 percentage points higher likelihood of exit, or 8% of the mean. By age 10, the difference is a 28% increase over the mean. Importantly, since the outcome measure is cumulative, the positive effect reflects the total success up to that year,

including the negative impact of the early years. Differentiation predicts higher outcomes over the firm lifecycle, but it takes time for them to occur.

Columns (2) through (4) disaggregate our equity growth variable into different types of exits. Column (2) considers only IPOs. We observe the same pattern, though the coefficient is significant until age 9. By age 10, moving from the 10th to the 90th percentile increases IPO probability by 33% relative to the mean. Column (3) is acquisitions. This pattern is similar but the coefficient turns statistically significant from age 6. By age 10, a startup at the 90th percentile is 27% more likely to be acquired than a startup at the 10th percentile. Finally, column (4) focuses only on high-value acquisitions (i.e. over $100 million). Interestingly, the coefficient, in this case, is not negative in the early years, but it does become positive later on. By age 10 it represents an increase of 43% relative to the mean. These results provide further robustness validating the dynamic effects evidenced in column (1). The evidence is consistent with our differentiation score ultimately capturing the nature of more innovative or unique ideas, which perform better over the long term, but they take a longer time to achieve this performance, possibly due to the cost of educating and better understanding the market to either acquire legitimacy (Deephouse 1999) or to prove technological feasibility (Marx et al. 2014).

How much does founding strategy matter?

Finally, we study the extent to which our measures, and consequently founding strategy, are economically relevant. To do so, we perform an out-of-sample analysis to consider how much variation in outcomes can be predicted by our measures. Specifically, using a 10-fold approach, we regress a logit model with a fully interacted version of our four founding differentiation scores (5 closest public firms, closest public firm, 5 closest cohort startups, and closest cohort startup) on the binary version of both of our outcomes, *Gets Early Stage Financing* and *Equity Growth*. We then store the out-of-sample predictions from these models,[24] and study how well do these out-of-sample predictions relate to realized outcomes. The results are reported in Figure 4.

[24]In essence, we split the data into 10 random subsamples, and for each subsample, we use the predicted value from a regression using all other 9 subsamples but excluding the focal one.

Panels A and B consider early stage financing. Panel A reports the share of firms that receive early stage financing across the distribution of out-of-sample predicted probability. We observe a positive and increasing slope, with firms at the top end of the distribution being about twice as likely to get early stage financing as firms at the bottom. Panel B is our preferred measure. It reports the out-of-sample ROC score (area under the curve) of this model. This is an established approach to assess the predictive fit of binary models. The ROC score conceptually answers the following question: if two firms, one with early stage financing and one without, are fed to the model, what is the likelihood that the one with early stage financing is scored higher by the model? A random model would have an ROC of 0.5, and a fully informative one would be 1. The graph shows an ROC value of 0.65, implying the model is able to account for about 30% of the variation in outcomes.

Panels C and D consider the same two statistics for the equity growth outcome. Once again we observe a meaningful ability of our index to predict performance. Firms in the top ventiles of the out-of-sample predicted index are about three times more likely to have an equity growth outcome than firms in the bottom ventiles. The ROC score is slightly lower at 0.60. Our model is able to account for 20% of variation in outcomes.

Appendix Figure A3 repeats the model using a simpler measure of similarly—the cosine of the term frequency-inverse document frequency (tf-idf) estimates. Interestingly, these estimates are lower. The ROC score for early stage financing is 0.62, while the ROC for the equity growth outcome is only 0.516, implying the tf-idf measures only account for about 3.2% of total variation. Finally, Appendix Figure A4 reports a model that uses both HP industries and differentiation scores together. The ROC scores increase, but only moderately, to 0.68 for *Gets Venture Capital* and 0.63 for *Equity Growth*.

Together, these estimates provide a novel assessment of the importance of founding positioning on overall performance. We show that founding positioning accounts for 30% and 20% of variation in financing and equity outcomes, respectively, out-of-sample. Drawing a parallel to the theoretical concept that it is not only using unique resources but being able to use them together

in novel ways, that creates strong strategic positioning, the tf-idf model, which does not consider the context in which words are used, scores much lower than our doc2vec model, which does. Furthermore, because our measures are inherently noisy, our estimate on the economic importance of founding positioning is possibly a lower bound within our sample.

5 Conclusion

Building on existing work using text-based machine learning, we have developed a novel approach to measure the strategic differentiation of startups and validated its role in predicting startup performance. Our approach focuses on the idea that companies state what their main value proposition is in marketing materials, and that websites are core marketing channels for most firms. A historical version of these websites, paired with natural language processing methods, can allow measuring distance in the website text from one company to another at specific points in time, and this distance can be aggregated into a measure of market differentiation from a company's closest competitors. We implemented our approach on data from Crunchbase to measure startup founding differentiation from its closests incumbent firms. We show that founding differentiation predicts financing and long-term performance, and accounts for a meaningful portion of the variation in outcomes. Founding strategy matters.

Our paper is designed to enable further research in this area. To do so, we have included several data and code appendices that allow incorporating our dataset and approach into other contexts. Our paper is accompanied by the release of four distinct datasets preserved in the Harvard Dataverse: i) a dataset containing all website text downloaded for the specific companies in our sample, including individual snapshots of each public company by year; ii) a dataset containing the doc2vec models estimated through this data for each year in the sample, and the Hoberg & Phillips tf-idf model using all websites; iii) a dataset using these models to estimate a matrix of text-based similarity between startups and other companies in the data each year; iv) a dataset including the estimated strategic differentiation score for each company. Our purpose in releasing this is to allow any researcher to take advantage of the data we have provided and expand the

knowledge around measuring and understanding startup strategy.

This data release is also accompanied by two distinct pieces of code, including i) an open source release of the code we use to scrape the Wayback Machine in Github and build our doc2vec models; and ii) a release of the regressions and code we have run to estimate all the tables in our paper. Due to sharing restrictions with Crunchbase, we are not able to release the main analysis data that includes Crunchbase private information, which was given to us through an academic license.

Building from these products, we hope researchers can continue expanding the analysis of text data, website information, and the way it predicts performance. Some of these avenues can include improving the NLP approach further by tuning specific parameters of the doc2vec model or introducing other, more advanced NLP models such as BERT; adding ancillary data beyond financing and equity outcomes, to understand other facets of firms and their evolution, such as patents, or workforce information; and expanding our approach beyond our sample to other contexts or countries. Whether, when, and how can strategy be measured is a rich avenue for future work.

Moving to entrepreneurial strategy, we hope our work is able to bring back emphasis on the relative importance of founding positioning for startups and their success. While strategy researchers had once pitted execution, dynamic capabilities, and founding positioning against each other as sources of competitive advantage, we are encouraged by how recent research instead recognizes the relative merits of each in startup performance (Koning et al. 2019, Gans et al. 2019). We provided an original estimate of how and how much does one facet of this equation matter, but we hope future work can continue applying analytical methods (such as machine learning) to these ideas to better elucidate what elements should go into an startup's ideal strategy.

Finally, considering our overall approach more broadly, we recognize that the role of data science and the IT revolution in shaping the way management research and practice is done is only starting to take shape. We have focused here specifically on using data science to measure a traditional construct in a new way, but other applications can include causal inference using ob-

servational data, and developing a new understanding of the key elements of strategic advantage that emerges from data-driven results, rather than using data to map existing theory. Ultimately, a data-driven approach to strategy may be quite distinct from the canonical, framework-based approach. We look forward to continued work in this space.

References

Abrahamson, E. & Hambrick, D. C. (1997), 'Attentional homogeneity in industries: the effect of discretion', *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* **18**(S1), 513–532.

Boulland, R., Bourveau, T. & Breuer, M. (2019), 'Corporate websites: A new measure of voluntary disclosure', *Available at SSRN 3816623* .

Bresnahan, T. F. & Reiss, P. C. (1991), 'Entry and competition in concentrated markets', *Journal of political economy* **99**(5), 977–1009.

Cennamo, C. & Santalo, J. (2013), 'Platform competition: Strategic trade-offs in platform markets', *Strategic management journal* **34**(11), 1331–1350.

Dalle, J.-M., den Besten, M. & Menon, C. (2017), Using crunchbase for economic and managerial research, Working paper, OECD Science, Technology and Industry Working Papers.

Deephouse, D. L. (1999), 'To be different, or to be the same? it'sa question (and theory) of strategic balance', *Strategic management journal* **20**(2), 147–166.

Eisenhardt, K. M. & Martin, J. A. (2000), 'Dynamic capabilities: what are they?', *Strategic management journal* **21**(10-11), 1105–1121.

Forbes (2020), 'Why every business needs a website'.
URL: *https://www.forbes.com/sites/theyec/2020/02/03/why-every-business-needs-a-website/?sh=610f549e6e75*

Gambardella, A., Camuffo, A., Cordova, A. & Spina, C. (2018), 'A scientific approach to entrepreneurial decision making: Evidence form a randomized control trial', *Management Science, Forthcoming* .

Gans, J., Scott, E. & Stern, S. (2019), *Entrpreneurial Strategy*, Manuscript.

Hoberg, G. & Phillips, G. (2016), 'Text-based network industries and endogenous product differentiation', *Journal of Political Economy* **124**(5), 1423–1465.

Igami, M. & Uetake, K. (2019), 'Mergers, innovation, and entry-exit dynamics: Consolidation of the hard disk drive industry, 1996-2016', *Available at SSRN 2585840* .

Koning, R., Hasan, S. & Chatterji, A. (2019), Experimentation and startup performance: Evidence from a/b testing, Working Paper 26278, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w26278*

Le, Q. & Mikolov, T. (2014), Distributed representations of sentences and documents, *in* 'International conference on machine learning', PMLR, pp. 1188–1196.

Marx, M., Gans, J. S. & Hsu, D. H. (2014), 'Dynamic commercialization strategies for disruptive technologies: Evidence from the speech recognition industry', *Management Science* **60**(12), 3103–3123.

McGahan, A. M. & Porter, M. E. (1997), 'How much does industry matter, really?', *Strategic management journal* **18**(S1), 15–30.

McGrath, R. G. & MacMillan, I. C. (2000), *The entrepreneurial mindset: Strategies for continuously creating opportunity in an age of uncertainty*, Vol. 284, Harvard Business Press.

Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013), 'Efficient estimation of word representations in vector space', *arXiv preprint arXiv:1301.3781* .

Mu, J., Bhat, S. & Viswanath, P. (2017), 'All-but-the-top: Simple and effective postprocessing for word representations', *arXiv preprint arXiv:1702.01417* .

Porter, M. E. (1996), 'What is strategy?', *Harvard Business Review* **6**(74), 61–78.

Quora (2013), 'How do i delete a company profile on crunchbase?'.
URL: *https://www.quora.com/How-do-I-delete-a-company-Profile-on-Crunchbase*

Reis, E. (2011), *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*, Currency.

Ruefli, T. W. & Wiggins, R. R. (2003), 'Industry, corporate, and segment effects and business performance: a non-parametric approach', *Strategic Management Journal* **24**(9), 861–879.

Rumelt, R. P. (1991), 'How much does industry matter?', *Strategic Management Journal* **12**(3), 167–185.
URL: *http://www.jstor.org/stable/2486591*

Schmalensee, R. (1985), 'Do markets differ much?', *The American economic review* **75**(3), 341–351.

Siggelkow, N. (2001), 'Change in the presence of fit: The rise, the fall, and the renaissance of liz claiborne', *Academy of Management journal* **44**(4), 838–857.

Teece, D. J., Pisano, G. & Shuen, A. (1997), 'Dynamic capabilities and strategic management', *Strategic management journal* **18**(7), 509–533.

Van den Steen, E. (2016), 'A formal theory of strategy', *Management Science* **63**(8), 2616–2636.

Table 1: Summary Statistics Crunchbase Firms

	mean	sd	min	max
Website Text Length	11227.623	16377	100	99668
Early Stage Financing (Thousands $)	909.157	3344	0	191000
Gets Early Stage Financing	0.682	.466	0	1
Series A Financing (Thousands $)	2113.331	5849	0	100000
Gets Series A	0.290	.454	0	1
IPO	0.017	.128	0	1
Acquisition	0.160	.367	0	1
High Value Acquisition (100M or more)	0.015	.12	0	1
Growth	0.177	.381	0	1
Observations	12406			

Dataset is all companies in Crunchbase founded since 2003 that raised financing and for whom we where able to download a founding website. Founding website is downloaded from the WaybackMachine as the earliest website available the year after founding. *Early Stage Financing* is defined as all financing that is seed financing, angel financing, or grants. *Website Text Length* is the number of total characters in the downloaded founding website text.

Table 2: Summary Statistics of Strategic Differentiation Score

	mean	sd	p10	p50	p90	min	max
Differentiation Score (5 Closest Public Firms)	0.635	.0636	.54	.64	.71	.46	.74
Differentiation Score (5 Closest Cohort Startups)	0.654	.0635	.57	.66	.73	.32	.85
Differentiation Score (Closest Public Firm)	0.600	.0722	.5	.61	.69	.11	.73
Differentiation Score (Closest Cohort Startups)	0.610	.0771	.51	.62	.7	.052	.83
Observations	12406						

Strategic differentiation score represents the conceptual distance in the market between a firm and some of its closest competitors. It is estimated in three steps. First, a measure of similarity is estimated between the founding website of all startups in a cohort and the website of all public firms during the startup year of founding. To do so, we use a word embeddings algorithm that accounts for both the incidence of words and their context. Next, distance is defined as one minus this similarity. Finally, differentiation is the average distance to the closest competitors. We report four measures. The distance to the five closest incumbent firms (public firms). Distance to the single closest public firm. Distance to the five closest startups from the same cohort. And distance to the single closest startup in the same cohort. All code is available at https://github.com/jorgeguzmanecon/measuring-founding-strategy.

Table 3: Does founding differentiation predict the receipt early stage financing?

	(1)	(2)	(3)	(4)
Differentiation Score (5 Closest Public Firms)	1.614***	0.664***	0.402***	0.372***
	(0.143)	(0.141)	(0.0624)	(0.0529)
Founding Year F.E.	No	Yes	Yes	Yes
HP Industry F.E.	No	No	Yes	Yes
Year × State F.E.	No	No	No	Yes
City F.E.	No	No	No	Yes
Observations	12406	12406	12406	12406
R^2	0.049	0.139	0.198	0.352

OLS linear probability model. Dependent variable is equal to 1 if a startup gets early stage financing (seed or angel financing) and zero otherwise. HP Industry fixed effects are fixed effects for 300 industries created by replicating the text-based industry approach of Hoberg & Phillips (2016) within our website data. Standard errors double clustered by HP industry and state. Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

Table 4: Does founding differentation predict the amount of early stage financing?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dep. Var. Early Stage	*Dep. Var.* Early Stage	*Dep. Var.* Early Stage	*Dep. Var.* Early Stage	*Dep. Var.* Seed	*Dep. Var.* Grant	*Dep. Var.* Angel	*Dep. Var.* Series A + Early Stage	*Dep. Var.* Series A *Subsample* Firms without Early Stage
Differentiation Score (5 Closest Public Firms)	20.59***	7.934***	4.545***	4.185***	3.025***	0.809**	0.823	1.931	1.663
	(1.843)	(1.744)	(0.792)	(0.715)	(0.845)	(0.361)	(0.540)	(1.200)	(2.324)
Founding Year F.E.	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
HP Industry F.E.	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year by State F.E.	No	No	No	Yes	No	No	No	No	No
City F.E.	No	No	No	Yes	No	No	No	No	No
Observations	12406	12406	12406	12406	12406	12406	12406	12406	3946
R^2	0.044	0.132	0.184	0.331	0.179	0.058	0.041	0.044	0.123

OLS model. Dependent variable is the log of total fundraised in early stage financing plus 1. HP Industry fixed effects are fixed effects for 300 industries created by replicating the text-based industry approach of Hoberg & Phillips (2016) within our website data. Standard errors double clustered by HP industry and state. Column (7) is the Series A fundraising only for companies that did not raise early stage financing. Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

Table 5: Other measures of founding differentiation and early stage financing.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Differentiation Score (5 Closest Public Firms)	4.545***				6.661***	7.699***	
	(0.792)				(1.784)	(1.396)	
Differentiation Score (Closest Public Firm)		3.203***			-1.960		3.259***
		(0.712)			(1.627)		(0.800)
Differentiation Score (5 Closest Cohort Startups)			1.282*			-4.223***	
			(0.607)			(1.169)	
Differentiation Score (Closest Cohort Startups)				1.478*			-0.0925
				(0.592)			(0.649)
Founding Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HP Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12406	12406	12406	12406	12406	12406	12406
R^2	0.184	0.184	0.183	0.183	0.184	0.185	0.184

OLS model. Dependent variable is the log of total fundraised in early stage financing plus 1. HP Industry fixed effects are fixed effects for 300 industries created by replicating the text-based industry approach of Hoberg & Phillips (2016) within our website data. Standard errors double clustered by HP industry and state. Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

Table 6: Does founding differentiation predict equity performance?

	(1) *Dep. Var.* IPO or Acq.	(2) *Dep. Var.* IPO or Acq.	(3) *Dep. Var.* IPO or Acq.	(4) *Dep. Var.* IPO	(5) *Dep. Var.* Acquisition
Differentiation Score (5 Closest Public Firms)	-0.0961	-0.112**	-0.127**	-0.0512**	-0.0607
	(0.0811)	(0.0463)	(0.0456)	(0.0231)	(0.0416)
Founding Year F.E.	Yes	Yes	No	Yes	Yes
HP Industry F.E.	No	Yes	Yes	Yes	Yes
Year × State F.E.	No	No	Yes	No	No
City F.E.	No	No	Yes	No	No
Observations	12406	12406	12406	12406	12406
R^2	0.099	0.142	0.262	0.111	0.126

OLS model. Dependent variable is a binary variable equal to 1 if a firm is IPO or acquired and zero otherwise. HP Industry fixed effects are fixed effects for 300 industries created by replicating the text-based industry approach of Hoberg & Phillips (2016) within our website data. Standard errors double clustered by HP industry and state. Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

Table 7: Founding differentiation and equity performance

	(1) *Subsample:* Drop Firms Founded 2012 or Later	(2) *Dep. Var.* IPO or Acq. During First 5 Years	(3) *Dep. Var.* IPO or Acq. After First 5 Years	(4) *Dep. Var.* Log(Acq. Price)	(5) *Dep. Var.* High Value Acq.
Differentiation Score (5 Closest Public Firms)	0.0189	-0.125**	0.0639	5.981**	0.00767
	(0.0693)	(0.0375)	(0.0431)	(2.238)	(0.0148)
Founding Year F.E.	Yes	Yes	Yes	Yes	Yes
HP Industry F.E.	Yes	Yes	Yes	Yes	Yes
Observations	5559	12406	12406	374	12406
R^2	0.128	0.054	0.134	0.500	0.040

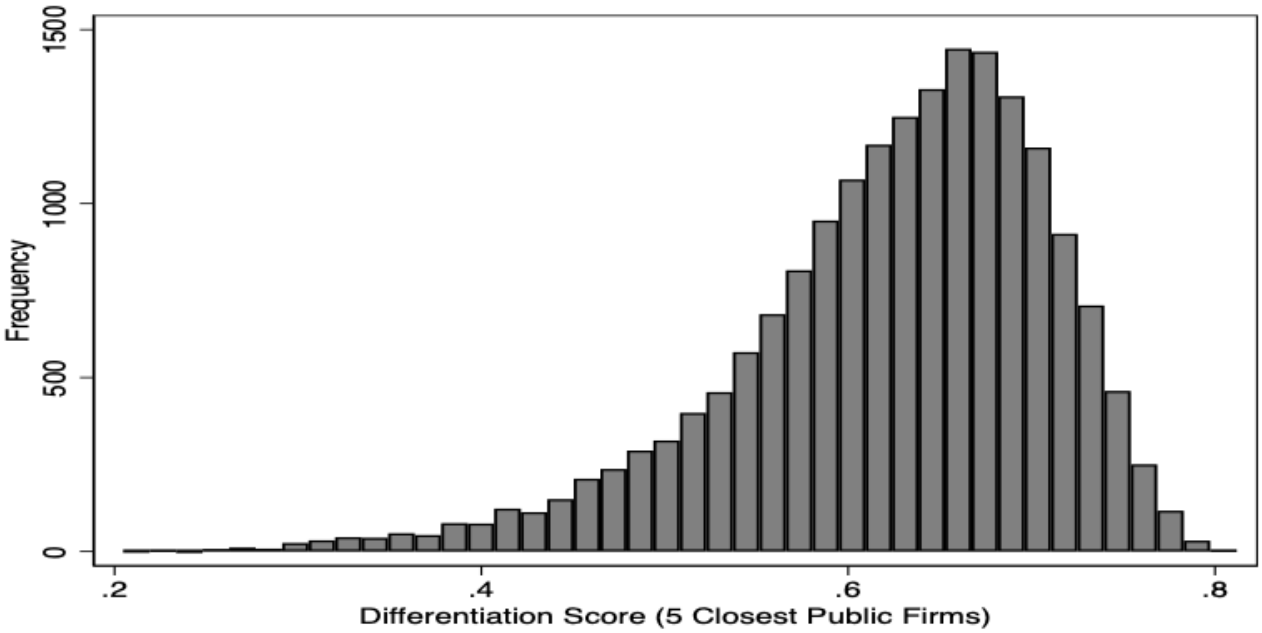
OLS model. Dependent variable is a binary variable equal to 1 if a firm is IPO or acquired and zero otherwise. HP Industry fixed effects are fixed effects for 300 industries created by replicating the text-based industry approach of Hoberg & Phillips (2016) within our website data. Standard errors double clustered by HP industry and state. Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

Table 8: Dynamic effects of founding differentiation on equity performance across age.

	(1) *Dep. Var.* IPO or Acq. (Cumulative)	(2) *Dep. Var.* IPO (Cumulative)	(3) *Dep. Var.* Acquisition (Cumulative)	(4) *Dep. Var.* High Value Acquisition (Cumulative)
Age=0 × Differentiation Score (5 Closest Public Firms)	-0.194**	-0.0195	-0.174**	-0.00517
	(0.0285)	(0.0165)	(0.0328)	(0.00776)
Age=1 × Differentiation Score (5 Closest Public Firms)	-0.182**	-0.0189	-0.163**	-0.00494
	(0.0282)	(0.0165)	(0.0314)	(0.00769)
Age=2 × Differentiation Score (5 Closest Public Firms)	-0.149**	-0.0169	-0.133**	-0.00363
	(0.0281)	(0.0166)	(0.0287)	(0.00751)
Age=3 × Differentiation Score (5 Closest Public Firms)	-0.103**	-0.0148	-0.0885**	-0.000371
	(0.0307)	(0.0164)	(0.0273)	(0.00691)
Age=4 × Differentiation Score (5 Closest Public Firms)	-0.0523	-0.0117	-0.0406	0.00297
	(0.0339)	(0.0164)	(0.0275)	(0.00645)
Age=5 × Differentiation Score (5 Closest Public Firms)	0.00282	-0.00916	0.0120	0.00822
	(0.0383)	(0.0160)	(0.0296)	(0.00607)
Age=6 × Differentiation Score (5 Closest Public Firms)	0.0566	-0.00163	0.0583*	0.0159**
	(0.0416)	(0.0170)	(0.0306)	(0.00585)
Age=7 × Differentiation Score (5 Closest Public Firms)	0.122**	0.00863	0.114**	0.0211**
	(0.0470)	(0.0187)	(0.0337)	(0.00722)
Age=8 × Differentiation Score (5 Closest Public Firms)	0.196**	0.0236	0.173**	0.0284**
	(0.0455)	(0.0207)	(0.0307)	(0.00847)
Age=9 × Differentiation Score (5 Closest Public Firms)	0.269**	0.0337*	0.236**	0.0407**
	(0.0444)	(0.0197)	(0.0316)	(0.00919)
Age=10 × Differentiation Score (5 Closest Public Firms)	0.346**	0.0434*	0.302**	0.0495**
	(0.0409)	(0.0227)	(0.0277)	(0.00854)
Age=11 × Differentiation Score (5 Closest Public Firms)	0.412**	0.0542**	0.358**	0.0542**
	(0.0396)	(0.0217)	(0.0254)	(0.00935)
Founding Year F.E.	Yes	Yes	Yes	Yes
Hp Industry F.E.	Yes	Yes	Yes	Yes
Observations	79711	79711	79711	79711
R^2	0.137	0.060	0.127	0.032

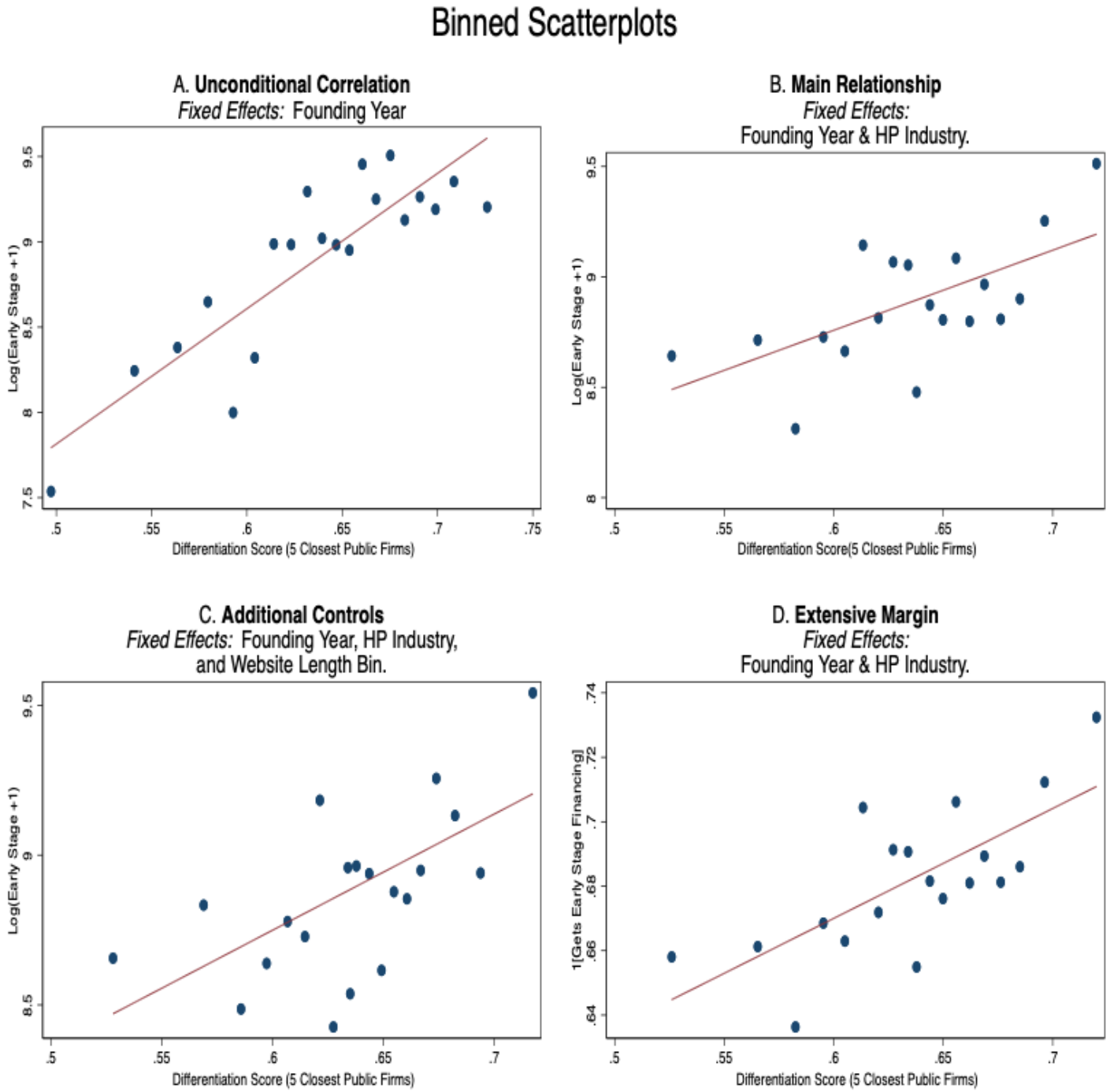
OLS model. Dependent variable is a binary variable equal to 1 if a firm has achieved IPO or acquired by age *t* and zero otherwise. Standard errors double clustered by HP industry and state. Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

Figure 1: Distribution of Strategic Differentiation Score (5 Closest Public Firms)



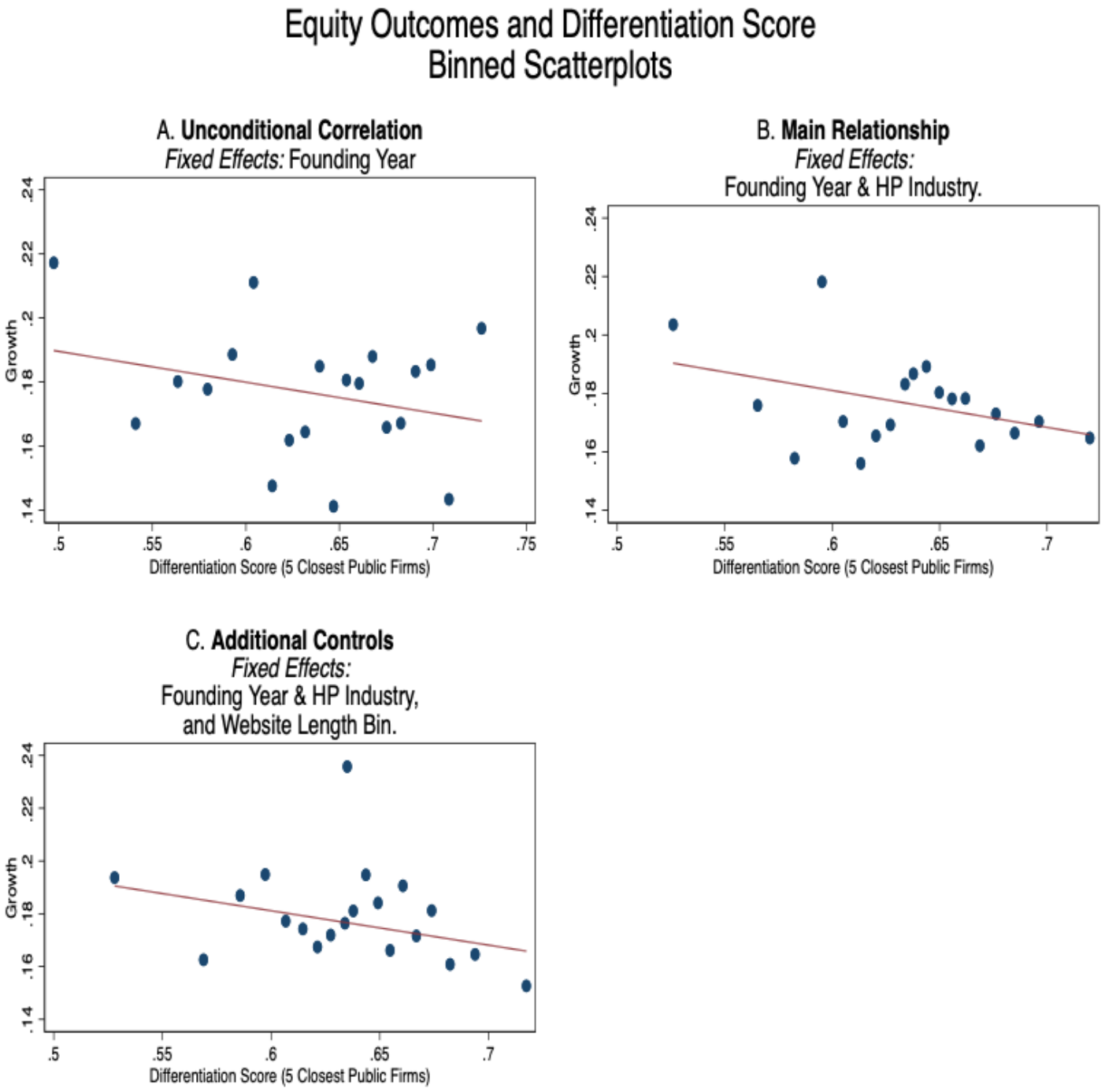
Notes: Reports the histogram of strategic differentiation score estimated as the mean distance in the founding website for the five closest public firms. Distance is one minus the similarity between websites, which is estimated using a word-embeddings algorithm of all public websites and startups in each cohort.

Figure 2: Differentiation Score and Early Stage Financing



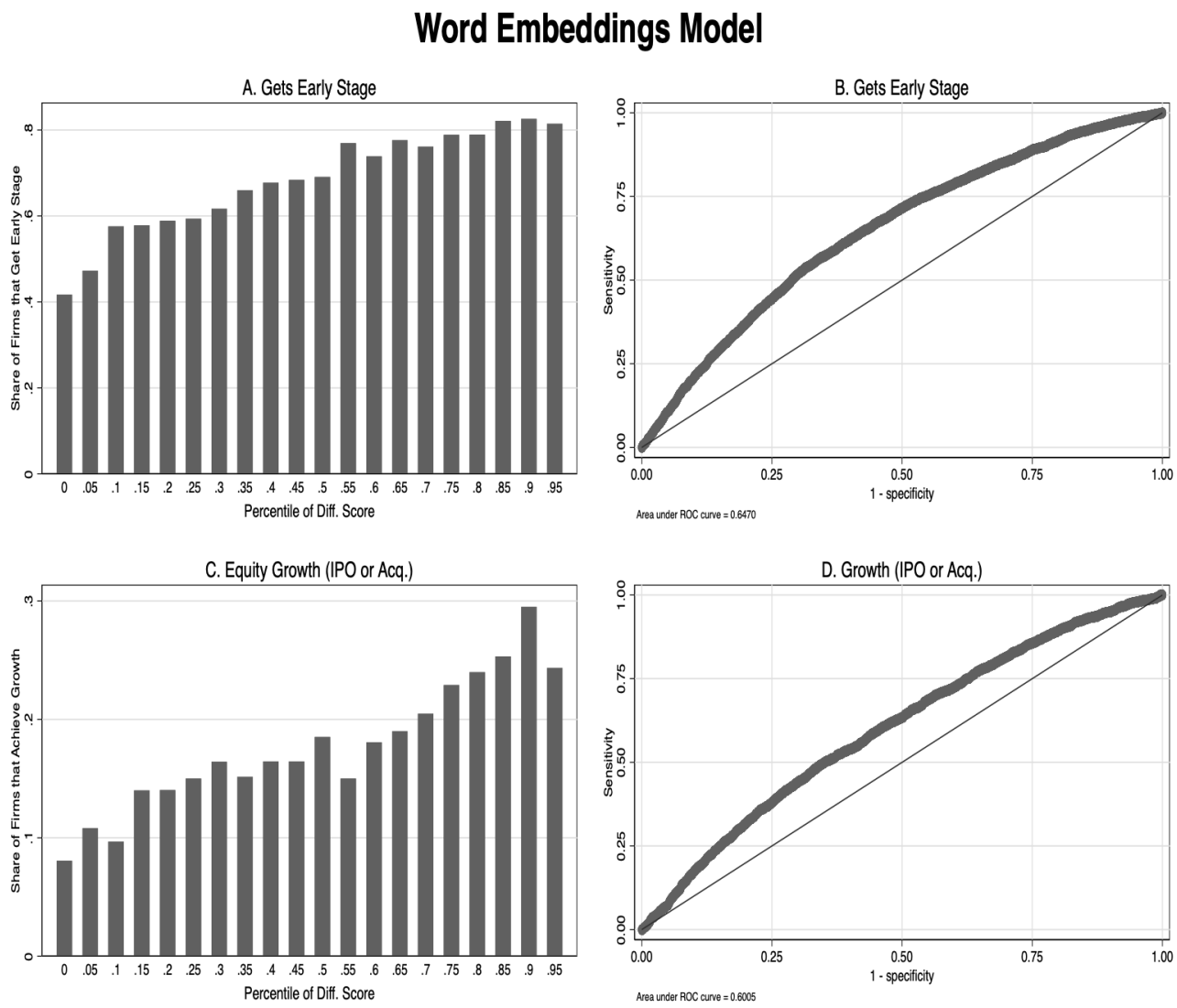
Notes: Early stage financing is all financing events recorded in Crunchbase as 'Seed', 'Angel', 'Crowdfunding', and 'Pre Seed'. HP Industry are the industries defined using the methodology of Hoberg and Phillips (2016) in our data. Appendix Figure A1 replicates these scatterplots with Series A financing events instead.

Figure 3: Differentiation Score and Equity Growth Outcomes



Notes: Growth is IPO or Acquisition. HP Industry are the industries defined using the methodology of Hoberg and Phillips (2016) in our data.

Figure 4: Out of Sample Predictability of Performance from Founding Text



Notes: This figure reports out of sample tests of how well do our measures predict performance. To do so, we run a fully interacted model of our four differentiation measures on two binary outcomes, *Gets Early Stage Financing* and *Equity Growth* using a 10-fold approach where we split the data into 10 groups and use the regression of 9 groups to predict the remaining one out of sample. Panels A and C report the distribution of outcomes across the predicted probability of performance. Panels B and D report the ROC (area under the curve) score which better measures the fit of the data.

Online Appendix to
Measuring Founding Strategy

Table A1: Correlation of Differentiation Scores

		(1)		
	(1)	(2)	(3)	(4)
(1) Differentiation Score (5 Closest Public Firms)	1.00			
(2) Differentiation Score (Closest Public Firm)	0.94	1.00		
(3) Differentiation Score (5 Closest Cohort Startups)	0.74	0.65	1.00	
(4) Differentiation Score (Closest Cohort Startups)	0.65	0.59	0.90	1.00

t statistics in parentheses

Figure A1: Other Measures and Early Stage Financing.

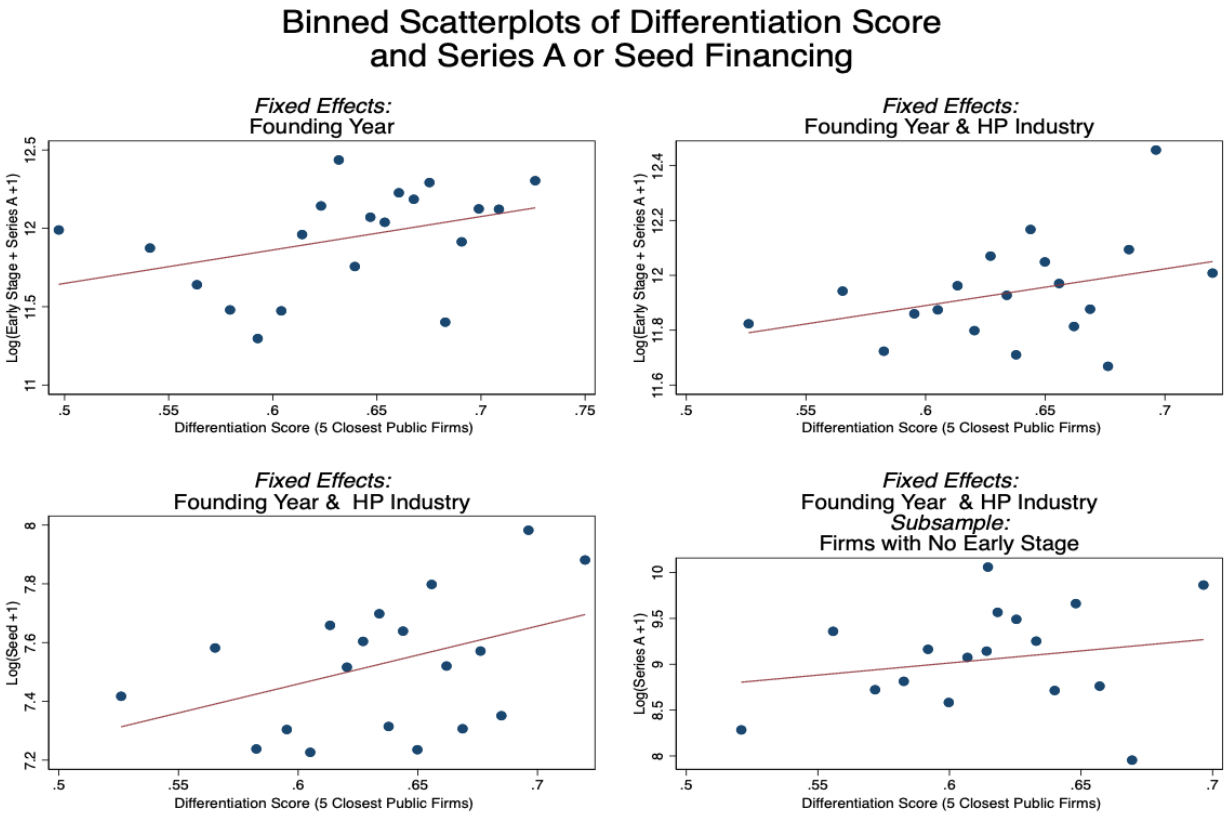


Table A2: Other measures of founding differentiation and equity performance.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Differentiation Score (5 Closest Public Firms)	-0.112**				-0.192	0.126*	
	(0.0463)				(0.155)	(0.0637)	
Differentiation Score (Closest Public Firm)		-0.0746			0.0738		0.0318
		(0.0543)			(0.160)		(0.0602)
Differentiation Score (5 Closest Cohort Startups)			-0.229**			-0.319**	
			(0.0530)			(0.0723)	
Differentiation Score (Closest Cohort Startups)				-0.163**			-0.179**
				(0.0431)			(0.0469)
Founding Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HP Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12395	12395	12395	12395	12395	12395	12395
R^2	0.141	0.141	0.142	0.141	0.141	0.142	0.142

OLS model. Dependent variable is a binary variable equal to 1 if a firm is IPO or acquired and zero otherwise. HP Industry fixed effects are fixed effects for 300 industries created by replicating the text-based industry approach of Hoberg & Phillips (2016) within our website data. Standard errors double clustered by HP industry and state. Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

Figure A2: Other Measures and Early Stage Financing.

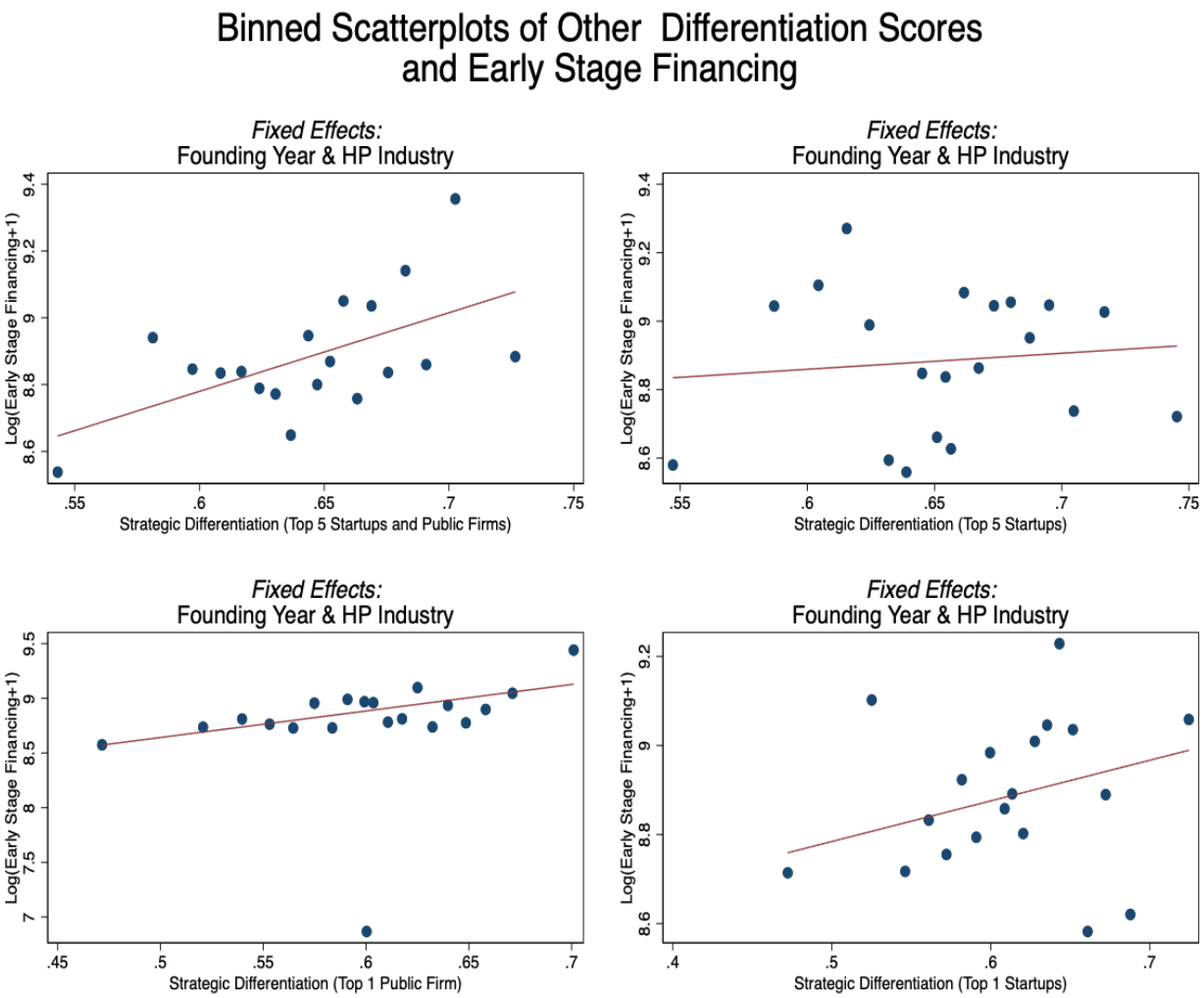


Figure A3

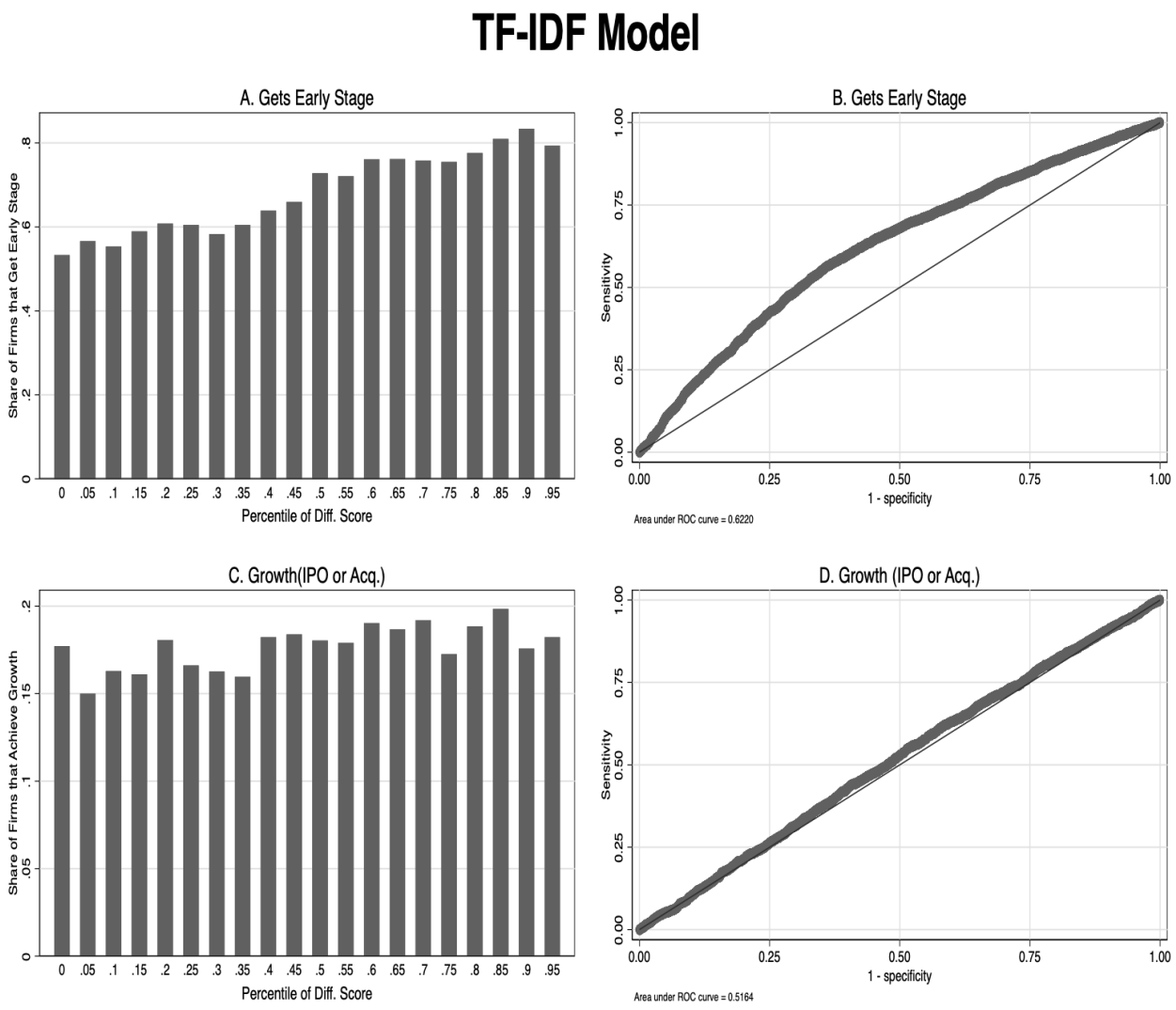

TF-IDF Model
A. Gets Early Stage
Share of Firms that Get Early Stage
0
.2
.4
.6
.8
0 .05 .1 .15 .2 .25 .3 .35 .4 .45 .5 .55 .6 .65 .7 .75 .8 .85 .9 .95
Percentile of Diff. Score
B. Gets Early Stage
Sensitivity
0.00
0.25
0.50
0.75
1.00
0.00 0.25 0.50 0.75 1.00
1 - specificity
Area under ROC curve = 0.6220
C. Growth(IPO or Acq.)
Share of Firms that Achieve Growth
0
.05
.1
.15
.2
0 .05 .1 .15 .2 .25 .3 .35 .4 .45 .5 .55 .6 .65 .7 .75 .8 .85 .9 .95
Percentile of Diff. Score
D. Growth (IPO or Acq.)
Sensitivity
0.00
0.25
0.50
0.75
1.00
0.00 0.25 0.50 0.75 1.00
1 - specificity
Area under ROC curve = 0.5164

Figure A4

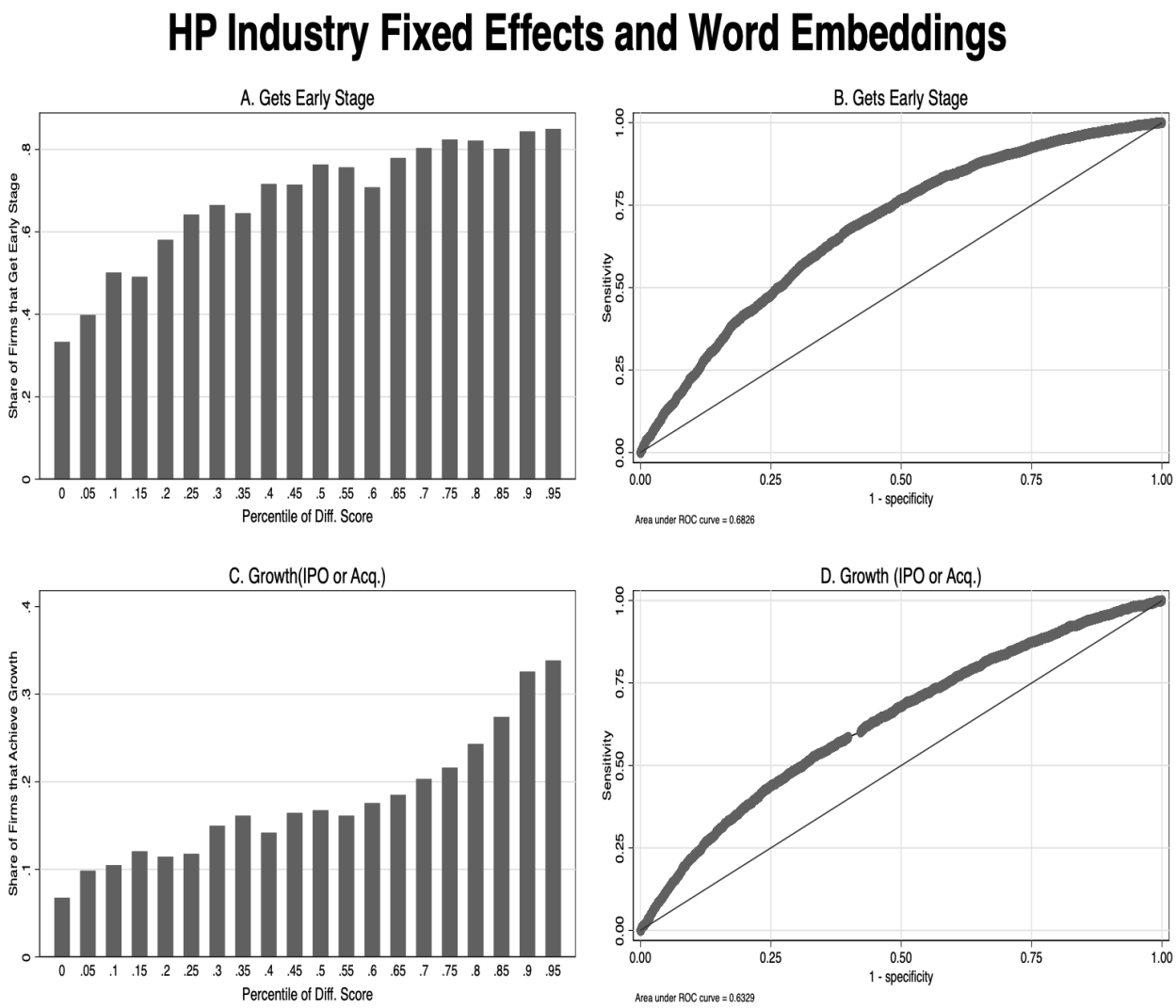

HP Industry Fixed Effects and Word Embeddings
A. Gets Early Stage
Share of Firms that Get Early Stage
0
.2
.4
.6
.8
0 .05 .1 .15 .2 .25 .3 .35 .4 .45 .5 .55 .6 .65 .7 .75 .8 .85 .9 .95
Percentile of Diff. Score
B. Gets Early Stage
Sensitivity
0.00
0.25
0.50
0.75
1.00
0.00 0.25 0.50 0.75 1.00
1 - specificity
Area under ROC curve = 0.6826
C. Growth(IPO or Acq.)
Share of Firms that Achieve Growth
0
.1
.2
.3
.4
0 .05 .1 .15 .2 .25 .3 .35 .4 .45 .5 .55 .6 .65 .7 .75 .8 .85 .9 .95
Percentile of Diff. Score
D. Growth (IPO or Acq.)
Sensitivity
0.00
0.25
0.50
0.75
1.00
0.00 0.25 0.50 0.75 1.00
1 - specificity
Area under ROC curve = 0.6329

Table B1: Examples of Strategic Differentiation Score: Consumer Electronics

Year	Company Name	Perc.	Diff. Score	Website	Short Description
2012	Penxy	99	0.735	penxy.com	Penxy is a slide sharing application for presenters to control their presentations in real time via iOS devices.
2013	Wallflower Labs	98	0.734	www.wallflower.com	Wallflower prevents home fires by monitoring your stove and alerting your smartphone if its on for too long or you forget to turn it off.
2016	Zoi, Inc.	98	0.733	www.whatiszoi.com	Software for wearables that will alert the user/family/first responders if they have abnormalities in their sleep.
2011	Healthy Stove, Inc.	98	0.732	healthystove.com	Developer of the world's first fully interactive modern microwave oven with a multi-touch screen interface, the NEXTWAVE smart oven.
2012	Crowd Supply acquired by Mouser Electronics	97	0.729	www.crowdsupply.com	Where engineers and designers crowdfund awesome new products.
2013	WearPoint	96	0.728	wearpoint.com	WearPoint is a developer and manufacturer of human input and wearable technology devices.
2013	BeON Home	96	0.727	beonhome.com	BeON Home, a better simpler way to provide safety and security in the home
2014	Slimcard	95	0.724	myslimcard.com	Slimcard helps you to get discounts worldwide, and will free up space in your wallet from unnecessary plastic cards.
2015	RC Ski	95	0.724	www.rcski.com	RC Ski develops remote control technology for personal watercrafts
2013	MobileQubes	94	0.722	www.mobilequbes.com	MobileQubes provides its clients with self-automated kiosks that allow customers to rent and return compact battery Qubes.
2011	TurtleCell	94	0.721	www.turtlecell.com	TurtleCell offers a thin protective case for iPhone and iPod Touch with built-in retractable earbuds.
2013	Wayv Technologies Ltd	93	0.720	www.wayvtech.com	WAYV products provide up to 30 minutes of instant and constant heat, whenever and wherever you may need it.
2014	FLUID	93	0.718	www.fluidwatermeter.com	FLUID - The Learning Water Meter used to conserve water, save money and detect leaks in real time.
2017	Amber Solutions, Inc.	92	0.716	www.ambersi.com	Ambers breakthrough technologies in advanced solid-state, IoT solutions brings a new standard to smart home and building infrastructure.
2012	ZON	91	0.716	zon-technology.com	Welcome to ZON, a technology solution company catering to the charging demands of mobile device consumers.
2012	CrossWorld Warranty	91	0.715	www.crosswarranty.com	CrossWorld Warranty is a platform for manufacturers and retail consumers to handle their warranty records online.
2013	Bluesmart	91	0.714	bluesmart.com	Bluesmart develops Internet connected travel products that include physical products combined with software.
2014	Biome	90	0.713	www.biome.us	The world's largest distributed wilderness.
2012	Plum	90	0.712	www.plumlife.com	The Plum Lightpad is the most advanced WiFi light dimmer available.
2014	Embr Labs	87	0.706	www.embrlabs.com	Embr Labs is a technology company that focuses on wellness that harnesses the power of cutting-edge technology to improve thermal comfort.
2013	Joylux	86	0.705	www.joyluxinc.com	Joylux creates innovative health solutions targeting the enormous.
2014	PogoTec	86	0.704	www.pogotec.com	PogoTec is the manufacturer of PogoCam. The worlds smallest, lightest camera that attaches to virtually all glasses.
2012	Qeexo	85	0.704	www.qeexo.com	Qeexo develops novel interaction techniques for mobile devices.
2012	August Home acquired by ASSA ABLOY Group	84	0.702	www.august.com	August develops smarter home access products and services including the award winning August Smart Lock.
2014	FlipFlic	83	0.700	www.flipflic.com	The easiest way to make window blinds smart.
2011	Anser Innovation	83	0.699	www.anserinnovation.com	Anser Innovation is pioneering interactive IoT technologies that enable remote care.

2013	Eyefluence acquired by Google	82	0.696	eyefluence.com	Eyefluence is the first eye-interaction technology for AR, VR, and MR devices which transforms intent into action through the eyes.
2014	Nucleus	81	0.695	www.nucleuslife.com	Nucleus is a wireless home intercom system that can be activated with voice.
2015	Voltus	80	0.694	www.getvoltus.com	Mobile power + expansion for your MacBook
2014	Need Fixed Parts	80	0.694	www.needfixedparts.com	Providing wholesale mobile device repair parts to repair shops across the United States
2017	Misty Robotics	77	0.689	www.mistyrobotics.com	Misty Robotics, a spinoff of Sphero, is a hardware company that builds personal robots for homes and offices.
2016	Purple Technologies	77	0.689	www.welcometopurple.com	Purple Technologies creates beautiful products that enhance the human experience.
2011	WISETIVI	77	0.689	wisetivi.com	WISE TIVI is an online platform that develops an alternative for smart TV.
2012	Stickerlight	76	0.688	stickerlight.com	Stickerlight, a consumer electronics platform, allows its users to add illuminative, dynamic, and customizable branding to their laptop.
2015	Siren	76	0.687	siren.care	Siren Care creates Neurofabrics; first product is Siren Diabetic Socks that helps people living w diabetes prevent amputations.
2014	SpectralCast	75	0.685	www.spectralcast.com	SpectralCast designs and delivers specialty electronics that create an Organic Network.
2010	FINsix Corporation	74	0.684	www.finsix.com	The world leader in VHF power. Building small, light power systems with best-in-class performance.
2014	Echodyne	74	0.684	echodyne.com	Echodyne develops and manufactures meta-materials-based radar technologies.
2010	LX Enterprises	74	0.683	lxenterprises.com	LX Enterprises is an e-commerce platform offering household furniture and appliances.
2014	zGlue	74	0.683	www.zglue.com	zGlue is a Silicon Valley company offering a platform and process for building custom chips on demand.
2011	Lark	73	0.682	www.lark.com	Lark is the leading digital health company using AI & clinical science to deliver scalable, positive health outcomes in chronic disease.
2016	Parihug	72	0.681	parihug.com	Parihug makes pairable, wifi-enabled teddy bears that let you hug someone from anywhere in the world.
2013	Korner	72	0.681	www.kornersafe.com	Korner manufactures home security systems for renters and middle-income homeowners.
2014	Oco cameras	70	0.677	getoco.com	Oco Cameras offers a range of Wi-Fi cameras for homes and businesses.
2009	Werkadoo	69	0.676	www.werkadoo.com	Werkadoo is an online portal that matches projects with candidates based on behavioral traits and soft characteristics.
2015	Skarp Technologies	69	0.675	www.skarptechnologies.com	Skarp Technologies is a start up that has developed the first personal razor powered by laser technology.
2011	ivee	68	0.674	www.helloivee.com	Ivee is a Wi-Fi voice-activated assistant that can answer questions, respond to commands, and connect to many smart home devices.
2013	UpDroid	68	0.674	alphalem.com	UpDroid is a technology startup with the vision to jump-start the personal robotics industry.
2014	Tiger Eye Sensor, Inc.	68	0.674	www.tigereyesensor.com	Tiger Eye Sensor, Inc. is the leader in the hands-free, wearable, personal security device market.
2012	PiperScout	68	0.673	piperscout.com	PiperScout is a mobile solutions company that provides mobile accessories and managed mobility services to Consumers and businesses.
2018	Kangaroo	67	0.673	heykangaroo.com	Kangaroo is an electronics company that provides simple home security systems.
2015	WavCatcher, Inc.	66	0.671	wavcatcher.com	WavCatcher provides better signals to improve the mobile experience.
2014	Flip Technologies	66	0.671	www.fliptechnologies.in	Flip Technologies focused on building transportation and navigation technologies for the automobiles of the future.
2014	Panorics	65	0.669	www.panorics.com	Panorics develops fully immersive 360-degree video technology and products.
2015	Findster	65	0.669	www.getfindster.com	Your pet's location and well-being with you 24/7

2016	CatFi	64	0.668	catfi.com	The smartest feeder with cat-facial recognition technology.
2013	Halo Neuroscience	64	0.668	haloneuro.com	Halo Neuroscience is a neurotech company focused on using electrical neurostimulation to unlock human performance and treat human disease
2013	Concepter	64	0.667	concepter.co	Concepter - a startup studio, that creates award-winning products and a hardware accelerator with $100k fund.
2012	CIMCON Lighting	64	0.667	www.cimconlighting.com	CIMCON Lighting offers intelligent, software-based streetlight control solutions that help cities.
2013	Alces Technology acquired by Mantis Vision	64	0.667	alcestech.com	Alces Technology develops industry-leading 3D cameras for AR/VR, 3D scanning and security applications.
2015	Atmotube	63	0.666	atmotube.com	Atmotube - First truly personal, most affordable and most compact air quality tracker.
2008	ecoATM acquired by Outerwall	63	0.666	www.ecoatm.com	ecoATM is an e-waste recycling company. ecoATM kiosks evaluate and purchase used electronics from consumers for cash or store credit.
2007	Instinctiv acquired by SoundCloud	62	0.665	www.instinctiv.com	Instinctiv is a music management and discovery app that syncs a users entire music library to any desktop computer or mobile phone.
2014	Eight Sleep	62	0.665	eightsleep.com	The world's first sleep fitness company.
2012	Reelsonar	62	0.665	reelsonar.com	Reelsonar is a digital fishing equipments manufacturing company.
2008	Alsentis	62	0.665	alsentis.com	The AlSentis technical team has more than 25 years in thin-film sensing technologies.
2013	Lumoid	62	0.664	lumoid.com	Lumoid is a try before you buy service for consumer electronics.
2010	QFO Labs	61	0.663	qfolabs.com	QFO Labs is dedicated to creating new high tech gaming concepts such as the Quad Fighter and Mimix products.
2016	Lemon	60	0.662	www.lemon-california.com	Lemon is a corporation that designs, develops, and sells solar powered and audio consumer electronics.
2015	Magnivation	60	0.662	www.magnivation.com	Magnivation creates the Toys of The Future - Personalized 3D Printed Smart Toys that merge Imagination with Interactive Play Experiences.
2015	NeoSensory	60	0.661	neosensory.com	NeoSensory gives people new senses via haptic feedback.
2012	Bounce Imaging	59	0.661	bounceimaging.com	Throwable omnidirectional/360 cameras for first responders
2012	L8 SmartLight	59	0.659	www.l8smartlight.com	The L8, a device composed of 64 LED lights and a super LED light, communicates users' interests through light codes.
2005	Flipswap acquired by HYLA Mobile	58	0.658	www.flipswap.com	Flipswap operates an online trade-in platform for buying and selling used consumer electronics.
2014	GoPlug	58	0.658	goplugbags.com	Plug Bags designs and manufactures backpacks and roller cases with a built-in high power battery.
2011	Energy Harvesters LLC	57	0.656	www.energyharvesters.com	Energy Harvesters LLC is commercializing a personal electronics mobile power source called the Walking Charger.
2003	InvenSense acquired by TDK	56	0.656	www.invensense.com	InvenSenseprovides motion-tracking devices for consumer electronic products such as smartphones, tablets, game controllers, and others.
2016	iBeat	54	0.653	www.ibeat.com	iBeat is a breakthrough smartwatch that continually monitors users 24-7 heart activity and can get them immediate help in any emergency.
2015	TuringSense	54	0.652	www.turingsense.com	TuringSense is a Silicon Valley-based pioneer in wearable sports technology,
2013	Typo Keyboards	53	0.650	typokeyboards.com	Typo Keyboards is engaged in the development of a physical keyboard accessory for the iPhone.
2014	KitHub	52	0.648	kithub.cc	Creative and hands-on electronic kits.

2013	Light	51	0.646	light.co	Light combines revolutionary optics technology with an advanced image processing engine to make a camera that changes photography.
2013	LEVO Oil Infusion, Inc.	50	0.646	www.levooil.com	LEVO is a premium kitchen appliance for infusing oil and butter with herbs at the touch of a button.
2012	Stir	48	0.642	www.stirworks.com	Stir is a developer of smart kinetic desks for homes and commercial entities.
2010	Sparkbuy acquired by Google	48	0.642	sparkbuy.com	Sparkbuy is an online service that allows users to compare the prices and features of consumer electronics gadgets.
2012	Sproutel	48	0.642	www.sproutel.com	At Sproutel they make toys that help children newly diagnosed with a chronic illness learn and cope through play!
2009	KE2 Therm Solutions	46	0.638	ke2therm.com	KE2 Therm Solutions specializes in the development of energy saving electronic controllers for refrigerators and air conditioners.
2011	TriviaPad	45	0.637	triviapad.com	TriviaPad is an iPad-optimized, real-time, multiplayer trivia game.
2006	Quanlight	45	0.637	www.quanlight.com	Quanlight delivers lighting solutions such as projection light engines, signage, traffic signals, and theatrical lighting systems.
2014	Sereneti Kitchen	45	0.637	www.sereneti.com	Sereneti Kitchen aims to automate cooking in the home kitchen.
2014	Nikola	42	0.632	nikola.tech	Nikola offers an advantaged far-field technology that converts radio frequency (RF) energy into usable direct current power.
2009	Dropcam acquired by Nest Labs	42	0.631	www.dropcam.com	Dropcamdevelops a Wi-Fi camera with easy setupand continuous cloud recording.
2015	DashTag	38	0.624	www.getdashtag.com	DashTag is a Dutch/US sport company that developed the Dash.
2014	Cinemood	38	0.624	www.cinemood.com	CINEMOOD is a Russian home cinema startup.
2011	DEY Storage Systems	38	0.624	www.dey-sys.com	DEY Storage Systems is a storage virtualization software platform that delivers scalable self-service storage.
2013	Genetesis	38	0.623	www.genetesis.com	Genetesis has developed CardioFlux, the first tool to non-invasively generate images of the hearts electrical current distribution.
2004	United Keys	37	0.622	www.unitedkeys.com	United Keys engages in the development of technology for PC display input devices such as keyboards and keypads to private label customers.
2010	Ringz.TV	36	0.620	ringz.tv	Ringz, an addictive video sharing application, connects users to shared playlists that are watchable on any web-connected device.
2011	SOL REPUBLIC acquired by HoMedics	35	0.617	www.solrepublic.com	SOL REPUBLIC is a consumer electronics company developing and marketing headphones, collaborating with a community of musicians and fans.
2013	Entirely, Inc.	35	0.617	entire.ly	Building a new innovation-social
2004	Sling Media acquired by EchoStar	34	0.615	slingmedia.com	Sling Media is a technology company that provides multi-screen smart TV solutions for consumers and television operators.
2011	BitGym	32	0.611	bitgym.com	BitGym is a digital home gym for tablets and smartphones.
2009	Immedia acquired by Amazon	32	0.611	immediasemi.com	Immedia Semiconductor develops and markets semiconductor based ISP and video compression technology for consumer electronics.
2013	Pavlok	31	0.610	pavlok.com	Pavlok turns your resolutions into reality.
2008	Coveroo acquired by Zazzle	29	0.606	www.coveroo.com	Coveroo is a software, printing and brand licensing platform that enables consumers and partners to create highly customized products.
2012	BenMedica	29	0.604	www.benmedica.com	BenMedica helps insurers overcome the electronic channel to drive value at the point of care.

2012	Fosbury acquired by Verve	28	0.602	www.fosbury.co	Drive store traffic with Mobile Wallet marketing
2011	Guzu	27	0.601	www.guzu.com	Guzu provides turnkey electronics recycling solutions for consumers, businesses, and non-profit organizations.
2013	Gramovox	27	0.600	gramovox.com	Gramovox reimagines vintage A/V equipment into functional contemporary art.
2013	Asius Technologies	27	0.599	asiustechnologies.com	Asius Technologies develops, protects, and markets a system of in-ear technology through its product ADEL.
2014	Intellijoule	26	0.598	www.intellijoule.com	Intellijoule, Inc. is a less than two years old technology creator.
2003	WiQuest Communications	26	0.598	www.wiquest.com	WiQuest Communications develops ultrawideband solutions for PC, consumer electronics, and mobile systems companies.
2011	iBaby Labs	24	0.593	www.ibabylabs.com	Mobile Video & Baby Monitor with Social Networking
2013	Blueshift International Materials	19	0.579	blueshiftmaterials.com	We develop and integrate high performance materials for demanding applications
2004	Syndiant	16	0.569	www.syndiant.com	Syndiant is an optical device company providing light modulating panels for high-resolution small displays.
2016	Totemic	15	0.565	www.totemic.com	Totemic makes a device to monitor the health and safety of older adults in a completely passive way–with no wearables required.
2006	Beats Electronics acquired by HTC	14	0.562	beatsbydre.com	Beats Electronics offers premium consumer headphones, earphones and speakers, software technology and streaming music subscription services.
2005	Polyera	13	0.560	www.polyera.com	The Polyerasupplies semiconductor, dielectric and interfacial materials for the printed and flexible electronics industry.
2017	WOWCube (CubiOs Inc.)	11	0.552	wowcube.com	The Earth's First Twisty Game Console to hack human evolution by twisting and shaking
2006	mSilica acquired by Atmel	10	0.547	msilicaweb.com	Smart mixed-signal integrated circuits
2010	Wentworth Technology	10	0.547	wentworthtechnology.com	Wentworth Technology develops SpeedThru, a drive-thru headset system for the quick service restaurant market.
2012	Breathometer	7	0.528	www.breathometer.com	Breathometer develops a breath analysis technology that helps people make smarter decisions and improve their lives.
2009	Henge Docks acquired by Brydge Technologies	5	0.517	hengedocks.com	Henge Docks is a supplier of docking stations for Apple products.
2003	CorasWorks acquired by HumanTouch LLC	5	0.517	www.corasworks.net	CorasWorks Corporation designs and develops workplace application software on the Microsoft SharePoint platform.
2003	Skyhook	4	0.510	www.skyhookwireless.com	Skyhook is a worldwide leader in location, powering location and context for apps, adtech and devices.

Table B2: Examples of Strategic Differentiation Score: Food and Beverage Industry

Year	Company Name	Perc.	Diff. Score	Website	Short Description
2015	Coda Signature	98	0.732	www.codasignature.com	Coda Signature is an award winning infused products company based in Colorado.
2017	Ripe.io	97	0.731	ripe.io	Ripe.io is creating the Blockchain of Food.
2015	Nix86	96	0.727	nix86.com	Nix86 is the best way to manage supplier relationships and keep product costs under control in your restaurant.
2015	Generosity Beverages	95	0.725	generositywater.com	It produces highest quality alkaline water with a perfect pH balance .
2016	Bonumose	92	0.717	bonumose.com	Bonumose mission is goodsugar to good people at good prices.
2013	Nima	92	0.717	www.nimasensor.com	Nima is a science-driven company creating connected food sensors for peace of mind at meal time.
2011	Coravin	91	0.715	coravin.com	Coravin offers a proprietary wine access technology that does not disturb a wine's aging process even after the cork is removed.
2016	FanFood	89	0.710	www.fanfoodapp.com	Order your favorite concessions directly from your smartphone.
2017	BRANDED STRATEGIC ADVISORS	87	0.707	brandedstrategic.com	HOSPITALITY TECH, FOOD AND BEVERAGE INVESTMENT ADVISORY FIRM
2013	Bevi	86	0.706	bevi.co	Bevi is a smart water dispenser that customizes flavored and sparkling beverages on demand.
2013	Summit Wine Tastings	86	0.705	summitwinetastings.com	Founded in 2010, Summit Wine Tastings, LLC is a wine and spirits marketing and promotions company based in Chicago, Illinois.
2012	Altar	81	0.696	altarco.com	Altar offers Herbal & Botanical Mood Mixer, which can be enjoyed without alcohol or mixed with fine spirits.
2012	Good Day Chocolate	80	0.694	gooddaychocolate.com	Good Day Chocolate makes chocolate with benefits for young working adults.
2015	Oak Analytics	77	0.688	www.oakanalytics.com	Oak Analytics is a data science solutions for instant authentication.
2015	Actionable Quality Assurance	70	0.678	www.actionableqa.com	Simplify the task of Quality.
2014	Appetizr	69	0.676	helloappetizr.com	Appetizr is the first app that recommends what to eat by learning what you like to eat.
2015	Fuego Box	67	0.673	fuegobox.co	Fuego Box is a hot sauce of the month club focused on craft & small-batch hot sauces.
2008	RuckPack	66	0.670	www.ruckpack.com	RuckPack is a provider of a caffeine-free, two-ounce, blood orange energy drink.
2010	Rip van Wafels	65	0.669	www.ripvanwafels.com	Rip van Wafels aims to revolutionize the way Americans enjoy their daily coffee.
2013	American Ostrich Farms	59	0.661	americanostrichfarms.com	A vertically integrated company from genetics to branded CPGs. Based on the most sustainable red meat animal in the world - ostrich!
2015	Vibe, LLC	59	0.660	www.glodrinks.com	Glo is the first liquid activated, lighted drink infuser and is perfect for any occasion.
2004	Charles Chocolates	56	0.655	charleschocolates.com	Charles Chocolates Artisanal chocolate company in San Francisco.
2016	Banyan Technologies, Inc.	55	0.655	www.banyansource.com	B2B local food e-commerce platform
2011	Poached Jobs	55	0.654	poachedjobs.com	Poachedjobs.com believes that hiring and being hired in the food and drink industry doesn't have to be a life sucking proposition
2012	Ubrlocal	54	0.652	ubrlocal.com	To provide the technical solutions that the urban food movement needs in order to thrive.
2015	Two Track Malting	53	0.650	twotrackmalting.com	Two Track Malting grows a variety of grains on their 6th generation family farm.

2009	Farmigo	52	0.649	www.farmigo.com	Farmigo makes it possible to order fresh, local, farm-to-table food by creating a scalable alternative to the traditional supermarket
2015	The Food Corridor	51	0.648	www.thefoodcorridor.com	SaaS software for shared use kitchens.
2015	True Made Foods	51	0.647	www.truemadefoods.com	True Made Foods radically improved Ketchup, BBQ Sauce and Sriracha. Removing the corn syrup, reducing the sugar and adding veggies instead.
2011	Little Duck Organics	49	0.644	littleduckorganics.com	To provide kids with good food for a healthy lifestyle and world domination.
2015	One Kombucha	49	0.643	www.onekombucha.com	One Kombucha offers certified organic raw kombucha.
2014	FreshSurety	46	0.638	www.freshsurety.com	FreshSurety enables users to wirelessly assess fresh food for spoilage without having to manually disassemble their cartons.
2012	MATI Energy	45	0.638	matienergy.com	MATI Energy is redefining energy drinks for health and function driven consumers.
2017	Farm2Cook	44	0.635	www.farm2cook.com	Farm2Cook offers retail stores and home delivery service of organic farm products.
2015	theGOapp and GO SMART WiFi	43	0.633	www.thegoappworks.com	Smart data and re-targeting platform for the hospitality and the beverage industry
2015	Vinebox	42	0.631	www.getvinebox.com	Vinebox is a premium wine-by-the-glass subscription service.
2015	Yourder	41	0.630	www.yourdernow.com	Yourder is a smartphone dining app that enables users to order food from their favorite restaurants using their phones.
2015	The Drop Wine	41	0.629	www.thedropwine.com	The Drop is disrupting the wine industry by combining quality with mobile convenience and creating a Millennial-targeted lifestyle brand.
2010	Lightside Games	39	0.626	lightsidegames.com	Lightside is a social gaming company creating Bible-based games that feature graphics, strong music and fun gameplay.
2015	Star Vodka USA	37	0.622	www.starvodkausa.com	Star Vodka USA is manufactures vodka from Non-GMO corn.
2009	Tessemae's All Natural	36	0.620	www.tessemaes.com	Salad dressings, Condiments and Marinades made by hand.
2016	Glimpse-BarFinder	34	0.616	glimpsebarfinder.com	Glimpse-BarFinder is a mobile platform for the nightlife industry allowing them to connect with their customers.
2014	Jarly.me	34	0.614	www.jarly.me	Jarly is a subscription service that offers sample boxes of fresh baked goods. All treats are made by local artisans and curated by us!
2014	Wilde Brands	30	0.608	www.wildebrands.com	Wilde Brands is a natural foods CPG that produces Chicken Chips, the first ever meat-based snack chip. Chicken Chips launched in July 2018.
2014	Naturi	30	0.608	www.naturi.com	Organic, grass-fed Greek yogurt company. Less sugar. More protein. Entirely honest.
2013	Sportwater Beverages	29	0.605	sportwaterbev.com	Sportwater Beverages provides electrolyte formulation drinks for athletes.
2013	Aspire Food Group	28	0.603	www.aspirefg.com	Aspire Food Group manufactures a variety of food products made from crickets.
2014	Eyewiz	27	0.600	www.eyewiz.com	Eyewiz links tourists with tour operators in an ideal manner, turning the touristic experience into a lifetime memory.
2014	TippleBox	26	0.598	www.tipplebox.com	TippleBox offers users a classy and small scale cocktail and drinking experience.
2009	Dryhootch	26	0.598	www.dryhootch.org	Dryhootch is a place where Veterans can gather informally in a coffee house; a safe, comfortable, drug-and-alcohol-free environment.
2010	Amorini Panini	24	0.593	www.amorinipanini.com	At the heart of it all there is the vision to create the first Panini brand in the United States.

2009	Better Bean acquired by The Hain Celestial Group	23	0.590	www.betterbeanco.com	Better Bean offers a locally-grown line of freshly made beans in recyclable containers.
2012	Rowdy Mermaid Kombucha	21	0.585	www.rowdymermaid.com	Rowdy Mermaid Kombucha is a craft brewery
2016	Lecker Labs Limited	20	0.581	www.leckerlabs.com	Lecker Labs Limited is developing a fully automatic yogurt maker Yomee.
2010	Nu-Tech Foods	20	0.580	nutechfood.com	Nu-Tech Foods develops and manufactures desserts.
2011	Dinner Lab	18	0.576	dinnerlab.com	Dinner Lab is a membership-based social dining experiment that unites undiscovered chefs with adventurous diners who are looking for
2008	True Drinks	18	0.575	truedrinks.com	True Drinks makers of AquaBall Naturally Flavored Water,.
2010	JuiceBoxJungle	17	0.573	www.juiceboxjungle.com	JuiceBoxJungle offers spirited commentary, expert views, confessions from parents, and quirky humor in various categories.
2012	EatWith acquired by VizEat	17	0.572	www.eatwith.com	EatWith is a community that invites people to dine in homes, connect with hosts, share stories and enjoy homemade cuisine.
2014	Oath Craft Pizza	15	0.568	www.oathpizza.com	Treat yourself to real food treated right. Certified Humane, sustainably sourced. Treat yourself to real food treated right.
2014	Maple acquired by Deliveroo	15	0.566	www.maple.com	Maple is setting a new standard in food delivery by pairing the best chefs in the city with top-flight technology and logistics
2012	Everly	15	0.566	goeverly.com	Everly is a 2014 american action thriller film directed by joe lynch and written by yale hannon based on a story by lynch and hannon.
2014	Green Blender	14	0.562	greenblender.com	GreenBlender is a smoothie delivery service.
2008	SanTsti	13	0.559	santasti.com	SanTasti develops, produces, markets and distributes nonalcoholic beverages.
2014	Blue Prairie	13	0.559	www.blueprairiebrands.com	Blue Prairie is putting the power of our natural dietary fiber, prebiotics, in your foods and beverages.
2015	Berri Pro	13	0.556	www.berripro.com	A Santa Monica startup company.
2015	BugEater Foods	12	0.555	www.bugeaterlabs.com	BugEater Foods is a product development company that makes cricket protein-based food products.
2016	Everbowl	9	0.540	www.everbowl.com	Everbowl is craft superfood. Everbowl specialize in create your own superfood bowls.
2010	JustFoodForDogs	9	0.539	justfoodfordogs.com	Nutritionally Balanced, 100% USDA Ingredients, No Preservatives, Vet Recommended & Always Hand-made in Small Batches.
2014	Caribe Juice	7	0.528	www.caribejuice.com	Caribe Juice, is banking on others loving the juice as much as he does.
2014	Miyokos Kitchen	6	0.523	miyokoskitchen.com	Miyokos Kitchen is a food production company that produces artisan cultured Vegan cheese and butter.
2010	Cocomama Foods	5	0.519	www.cocomamafoods.com	Cocamama is a goddess of health and happiness.
2015	New Barn	5	0.518	www.thenewbarn.com	The Leading National Organic Almond Milk Brand .
2010	Spindrift	5	0.518	spindriftfresh.com	Americas first sparkling beverages made with FRESH squeezed fruit.
2017	Siren Snacks	4	0.511	www.sirensnacks.com	Siren takes the foods and flavors people love and recreates them using clean, plant-based ingredients.
2011	Tava Indian Kitchen acquired by Curry Up Now	4	0.510	www.tavaindian.com	Premium Indian fast food restaurant

2012	Thrive Nutritious Ice Cream	4	0.509	thriveicecream.com	Thrive is Premium Ice Cream In Seven Delicious Flavors
2012	Punch Bowl Social	4	0.508	punchbowlsocial.com	Punch Bowl Social Food & Drink is a restaurant, bar, bowling and entertainment concept.The concept utilizes bowling and other amenities to
2004	NadaMoo	4	0.507	www.nadamoo.com	Dairy-free, yet creamy; fewer calories a
2015	Petit Pot	4	0.506	petitpot.com	A French inspired, next level quality dessert company .
2015	Tastd	3	0.502	www.tastdapp.com	Tastd is an app that makes personal restaurant recommendations accessible.
2016	Brami	3	0.501	www.bramibeans.com	Brami reprocesses an ancient Roman snack for modern consumers.
2014	MerryMint	3	0.501	www.merrymint.co	The best local food & drink delivered for office celebration.
2006	Attune Foods acquired by Post Holdings	2	0.495	www.attunefoods.com	Attune Foods offers probiotic wellness bars in various flavors such as chocolate, blueberry, vanilla, peanut butter, mango, lemon.
2006	Boomerang Pies	2	0.489	www.boomerangspies.com	Boomerang makes hand-held savory pies that are sold in both the fresh and frozen food isle of your local grocery store.
2012	Marlo's Bakeshop acquired by Angel miracle cookies	2	0.487	marlosbakeshop.com	Marlo's Bakeshop crafts wholesome indulgences that nurture our customers, strengthen our community and promote women's entrepreneurship.
2015	JicaChips	2	0.486	www.jicachips.com	JicaChips- World's 1st Jicama Chip! Less than 100 Calories Per Bag.
2012	Spot On Foods	2	0.486	www.spotonfoods.biz	Founder of Spot On Foods, a Seattle-based wholesale baking company.
2012	Saucey Sauce	1	0.483	getsauceynow.com	Vietnamese-inspired Dressings, Marinades, Cooking & Finishing Sauces and Ketchup
2011	ips All Natural	1	0.477	ipsallnatural.com	IPS All Natural is a manufacturer of healthy and tasty snacks.
2008	Olomomo Nut Company	0	0.470	olomomo.com	Olomomo Nut Company produces and sells a variety of nut products and snacks.
2016	Oroplata Resources, Inc.	0	0.462	www.lithiumore.net	Lithium-Ore Producer

Table B3: Five Closest Startups and Public Companies to **whatiszoi.com**

Type	Similarity	Website	Website Text (first 500 characters)
public firm	.344	www.gtxcorp.com	GTX Corp (GTXO) — With You Home Tracking Solutions Our Tracking Solutions Stand Alone Devices Track My Work Force GPS SmartSole Platform Licensing Code Amber Alert Tag About Us About Us Board of Directors Management Team Advisors Strategic Partners Brand Ambassadors Investors Investors Message from the CEO Press Press Blog GTX Video Media Gallery Contact Us Portal Login Open Menu Home Tracking Solutions Our Tracking Solutions Stand Alone Devices Track My Work Force GPS SmartSole Platform Licen
public firm	.304	www.iomagic.com	Mobile Optical Drives - IOMagic Categories Power Rechargeable USB Battery Universal Travel Adapters USB Solar Chargers USB Wall Chargers USB Car Chargers Storage Mobile Optical Drives External Desktop Optical Storage Internal Desktop Optical Storage Hard Drive Enclosures Cables Apple USB Cables Micro and Mini USB Cables Replacement Cables Accessories Audio Bluetooth Audio Bluetooth Keyboards Cell Phone Car Mounts Fans Lighting Mice Smartphone Photography Cases and Covers iPhone Cases Phone Cases S
public firm	.28	www.brookstone.com	Brookstone Store Locator image svg+xml Loading 0 FREE SHIPPING WITH $99+ USE CODE: Exclusions image svg+xml image svg+xml R Brookstone Search Catalog Search Log In Find a Store Find a Store Customer Care: Log In Favorite Products X Featured Featured Star Wars Limited Edition Ariana Grande Cat Ear Headphones Big Blue WiFi Speakers with Chromecast Flight Force Drones Mio SLICE Heart Rate and Activity Tracker Hamleys Quick View Star Wars Speeder Bike Quadcopter Drone by Propel Free standard shipp
public firm	.268	dickssportinggoods.com	Track Your Order Message Dialog Close Free Shipping Over $49 Details My Account Account Summary Sign Out Help Track Order Shop Departments Shop By Sport Shop By Sport Baseball Basketball Bikes and Cycling Bowling Boxing and MMA Camping and Hiking Cheerleading Cricket Exercise and Fitness Field Hockey Fishing Football Golf Gymnastics Hockey Hunting and Shooting Ice Skating Kayak and Paddle Lacrosse Outdoor Living Rec Room Games Rugby Running Skates and Scooters Snowboarding Soccer Softball Swimming Table Ten
public firm	.266	www.biozoom.net	:: Biozoom, :: ENGLISH — GERMAN — SPANISH biozoom About Us Technology Benefits Market Approach Contact ORDER NOW What is biozoom? biozoom is a handheld transdermal spectrometer for instantly detecting Mobile, accurate and easy to use, biozoom analyzes data critical to managing and improving a healthy Proven effective in clinical trials, biozoom s findings are secure and available via smartphone or Learn more The biozoom Advantage biozoom puts physiological diagnostics in the palm of the hand,
startup	.419	www.ibeat.com	FAQ - iBeat - The Heart Monitoring Smartwatch That Can Save Your Life Get Notified iBeat The Heart Monitor Smart Watch That Can Save Your Life Home Features FAQ About Us Contact xd7 Sign up for an early discount and get notified when iBeat is available Sign Up Sign up for an early discount and announcements: Sign Up Frequently Asked Questions FAQ When will the smartwatch be available, and how can I buy it? We will begin taking for the iBeat Life Monitor online in If you sign up with your email
startup	.303	www.wayogps.com	Support - Wayo Support - Wayo xd7 Home Blog GPS Tracker Support Wayo Home Blog GPS Tracker Support Sign In Support Your Name (required) Your Email (required) Subject Your Message We already are Wayo users! Wayo Facebook Twitter LinkedIn Instagram GPS Tracker Press Blog Wayo Team Support Investor Relations Shareholder Privacy Policy Terms of Use Download from AppStore Download from Google Play xa9 Wayo All rights reserved Developed by Doonamis GPS Tracker - Wayo GPS Tracker - Wayo xd7 Home
startup	.291	www.tempoiq.com	Platform Toggle navigation Platform Solutions Industrial Automation Medical Devices Oil and Gas Resource Management Smart Grid Solar Energy Wearables Internet of Things News Blog Login Contact Us Compose Your IoT Application Now Connect IQ Data IQ Analyze IQ View IQ Cloud IQ ViewIQ View Composer With View Composer, you can create realtime, IoT dashboards, application, and visualizations without shareable with your team, your company, and your Modular and Extensible Integrate realtime IoT data visual
startup	.283	cardiomo.com	Cardiomo wearable biosensor which saves lives Home About Us Contact Stay close to the heart of your Loved Ones Saves lives by 7 health Subscribe now to find out more about Cardiomo Subscribe Can you know that loneliness kill? Researches found that people who feel lonely have a % higher risk of having heart disease and being alone could increase the risk of premature death by %. Early symptoms of heart disorders are very hard to decode and interpret correctly without medical Its prevention and
startup	.265	www.haiku.ai	Haiku - design prototyping to production code React App’

Table B4: Five Closest Startups and Public Companies to **hengedocks.com**

Type	Similarity	Website	Website Text (first 500 characters)
public firm	.541	www.amx.com	QuickShip Solutions Total Environments Total EXPerience Kits EXPerience Kits AMXmeetingroom AMXhome Products TOTAL ENVIRONMENT CONTROL Environment Controllers Cat5 Presentation Switchers Environment Accessories ControlPads USER INTERFACES Touch Panels Keypads Remotes ControlPads Presentation Interfaces Environmental Controls Camera Controllers User Interface Accessories DEVICE CONTROL ControlPads Central Controllers Control System Accessories Network Communication Lighting Controls Environme
public firm	.489	www.gateway.com	Gateway Official Site: Notebooks, Laptops, Desktops, , Displays, Monitors, Accessories Main Content Site Navigation Footer Shop Notebooks Desktops Displays Accessories Deals Where to buy Upgrades and parts Computers Netbooks Notebooks LT Series from $299 EC Series from $399 NV Series from $529 P Series from $1,099 Desktops SX Series from $499 DX Series from $509 LX Series from $729 FX Series from $1,129 Gateway One from $719 Which one is right for me? Gaming lineup Where to buy Displays ” Wide
public firm	.477	www.compxnet.com	CompX National - Office Furniture Need CompX instruction sheets? Download: Instruction sheets Terms: Sale xb7 xb7 Purchase Home Search Sitemap Downloads CompX Security Products CompX National CompX Fort CompX Timberline CompX Chicago CompX eLockxae Stock Locks CompX National View National products Find a distributor Download National catalog Download National catalog images About CompX National Contact CompX National CompX National - Office Furniture Manufactured to Order - Applications Adv
public firm	.473	www.hauppauge.com	Hauppauge Computer Works Dual tuner high definition ATSC QAM digital TV plus analog TV for Windows XP and Vista! Two tuners with a hardware encoder for analog Now supported by BeyondTV and SageTV, plus Windows Media Center and Hauppauge s WinTV version 7 Dual ATSC NTSC QAM TV tuner Two analog hardware encoders plus dual ATSC QAM digital TV receivers, on one half height PCIe board! FM radio and dual IR blasters Great for Windows Media Center or the WinTV v7 : $149 WinTV and Orb: Watch TV on you
public firm	.463	www.bestbuy.com	Best Buy - Page Not Found Search Sorry, the page you requested was not The page may be temporarily unavailable or may no longer There are a few things you can try: Return to the Home Page View the Site Map Shop our Name Brands Shop our main categories: TV and Video Audio Car and GPS Cameras and Camcorders Computers Mobile Phones and Office Music, Movies and Books Video Games and Gadgets Home and Appliances Best Buy Skip Navigation My Account Order Status Customer Service Espaxf1ol Best Buy - Buyer Be Wee
startup	.442	www.liquidware.com	Liquidware : Shop SHOP APP STORE PROJECTS WIKI ASK! WHAT S NEW ALL CABLES COMMUNITY COMPONENTS ILLUMINATO X MACHINA KITS LEGACY HARDWARE MODULES Cart: View Cart or Checkout The Shop Shop for hardware modules and to build your own International shipping available! more at the blog Like Liquidware? Featured 6 inch Type USB Cable ($9.36) ButtonPad ($31.95) GamePack ($249.93+) Illuminato X Machina ($65.89+) Lithium Backpack ($31.90+) Mega Backpack ($34.63+) TouchShield Slide ($174.93) TouchShield
startup	.394	www.iruleathome.com	iRule - iPhone and iPod Touch remote Allowing you to control your home About FAQ Devices Support Gallery Contact About the iRule What The iRule is an application that transforms the iPhone iPod Touch into a universal remote The iRule allows you to control all of your audio video gear simply and The iRule focuses on a simple user experience that allows non experts the ability to quickly and easily create a powerful custom Upgradeable The iRule is a software that allows you to customize your inter
startup	.362	www.whatsapp.com	WhatsApp Faq I love this app!!! Hands down best text app ever! AirIsMusic (App Store US Review) Apple should ve thought of this This app must be on every iPhone home TheIslander (App Store US Review) This app is trully the best app ever made and it is getting (App Store US Review) Now the iPhone is complete! Store UK Review Home Download FAQ Blog Contact WhatsApp FAQ What is it? Can I also chat with my BlackBerry contacts?? How is WhatsApp Messenger different from SMS? Help, my iPhone Push No
startup	.333	gdgt.com	gdgt gdgt gdgt is the new consumer electronics site by Peter Rojas and Ryan Block – the guys behind Engadget and We re still prepping things (no, this isn t the final site!), but we ve got a weekly podcast you can listen to in the Feel free to leave your email address if you want to be notified when we launch in You can also reserve your user name in advance, but it ll be on a first come first serve (If someone else requested your name first or we just can t offer it, we ll let you know when
startup	.322	pocketgems.com	Pocket Gems - Welcome Games About Jobs Support Contact Easy to Pick Up, Impossible to Put Down Previous Next Pocket Gems is building the next generation of wildly engaging mobile Our games are , making them accessible to everyone, and pack the most fun per square inch on any gaming platform! We re hiring! Join us and revolutionize the market for mobile games xa9 Pocket Gems, All Rights Pocket Gems - Support Games About Jobs Support Contact Welcome to the Pocket Gems Support Page Below you w

Table B5: Five Closest Startups and Public Companies to **codasignature.com**

Type	Similarity	Website	Website Text (first 500 characters)
public firm	.27	www.publix.com	Welcome to Publix — Publix Super Markets Weekly Ad My Publix close Find a Store Back Advanced Search Enter Zip Code, City and State, or Store Number Search Locator service by Systems Use the search form on the left to find a My Shopping List close Whoops! We re this Something went wrong while we were trying to handle your Please try If you need assistance Customer Care or call us at (800) Your list is Get started now! Enter Items Browse Products Other ways to add items: Weekly Ad Recipes Featu
public firm	.27	www.pulsebeverage.com	Natural Cabanaxae Lemonades - Pulse Beverages SHARE Home Buy Natual Cabanaxae Lemonades Natural Cabanaxae Limeades Natural Cabanaxae Lemonade and Limeade Combination Natural Cabanaxae Coconut Waters PULSE Heart and Body Health About Nutritional Info News Releases PULSE Blog Privacy Policy Disclaimer Investors Contact Buy Natural Cabanaxae Lemonades Natural Cabanaxae Lemonades Natural Cabanaxae Blueberry Lemonade $26.95 $26.95 Natural Cabanaxae Cherry Lemonade $26.95 $26.95 Natural Cabana
public firm	.262	www.cariboucoffee.com	About Our Coffee Life is Stay awake for Home Shop Online About Our Coffee Menu and Nutrition Our Locations Caribou Cards Sign In Customer Service Your Account Shopping Cart About Our Coffee Our Coffees Rainforest Alliance Quality Advantage Brew Your Best FAQ Roastmasters Blog Coffee Recommender From Our Roastmasters Roastmaster s Blog Caribou s Proactive Partnership with The Rainforest Alliance February , Hacienda La Minita January , All Blog Posts Roastmaster s Report Winter Winter Featuring Pa
public firm	.249	www.panerabread.com	Panera Bread Menu and Nutrition Home Our Menu Our Paneraxae Catering The Panera Cardxae MyPanera Rewards About Us From the Bakery From the Cafe Nutritional Calculator Today s soups are Broccoli Cheddar, Chicken Noodle, French Onion, Cream of Chicken and Wild Rice and Vegetarian Black Additional soups are available Call your for their daily soup Download the entire Panera menu Menu and Nutrition With the skill of an artisan, the heat of the oven and a few fine ingredients, our bakers make bread th
public firm	.245	www.gordmans.com	Shipping Free In Store Returns! FREE SHIPPING on orders $75 or more! Details Save UP TO % OFF department store prices every day! My Bag (0) Find a Store Contact Us Save for Later Mobile Navigation Log In Rewards Gordmans Gordmans Search Catalog Search Skip to content Close Search Catalog Search Holiday Holiday Decor Decorative Accents Decorative Word Signs Figurines Holiday Florals Ornaments and Garlands Tree Skirts and Stockings Tabletop Trees Wall Art Holiday Entertaining Set the Table Bakeware
startup	.288	ascendcannabis.co	Where To Buy - Ascend Cannabis Skip to content Ascend Cannabis Ascend Premium Cannabis Oil + Terpenes Elevate your lifestyle Home Products Where To Buy Press Contact Us Home Products Where To Buy Press Contact Us OUR RETAIL DISPENSARY PARTNERS RETAILER? CLICK HERE Your location Search radius mi mi mi mi mi mi Results Copyright - All Rights Ascend Cannabis – Retailer? Click Here YOU MUST BE + TO ENTER By pressing enter, you confirm that you re +. Enter Exit cwv3_den_title cwv3_den_msg Exit A
startup	.253	www.jicachips.com	Jicama Nutrition and health benefits — JicaChips Cart Products Meet JicaChips Why Jicama? Our Story FAQ Blog Contact Menu Products Meet JicaChips Why Jicama? Our Story FAQ Blog Contact Cart Why Jicama? Want exclusive deals and healthy recipes? Links FAQ Store Locator Product Request Form Contact Privacy Follow Us xa9 JicaChips Want exclusive deals and healthy recipes? Jicama Chips JicaChips — JicaChips Cart Products Meet JicaChips Why Jicama? Our Story FAQ Blog Contact Menu Products Meet JicaCh
startup	.252	www.starvodkausa.com	Shop Accessories STAR VODKA Home About Purchase Star Vodka Shop Accessories Seasonal Potions Media Contact STAR VODKA Home About Purchase Star Vodka Shop Accessories Seasonal Potions Media Contact Shop Accessories Purchase all the latest Official Star Vodka products and accessories! FREE SHIPPING on orders over $75 use Pomo Code: I love free Badass Star Trucker Cap Sipping on Star Shots (2 pack) Hamptons Star Cap Skull Star Cap All Color Top Star Vodka USA All Rights Reserved About STAR VO
startup	.251	www.spinn.com	Home - Spinn CoffeeSpinn Coffee HOME ROASTERS TEAM KEEP UP WITH SPINN Join the Spinn Coffee community for your chance to win a Spinn coffee machine or coffee supply for a year! One free 1 lb of coffee each month for At we know your morning cup of coffee is much more than an early morning Its a ritual, a moment of tranquility, a daily treat for your taste Brewed in a new way without the use of cups, pods or THE BEST ROASTS ARE HERE JUST BEANS, NO PODS Spinn offers a wide variety of the best loc
startup	.246	petitpot.com	Petit Pot - Gourmet Desserts - Bonjour! Home Who We Are Desserts Pot de Creme Cookies Where News Contact At Petit Pot we bring a delightful blend of French pastry practice with modern California culinary Using simple, sustainable practices with local goods, were able to make the best We hope you enjoy a sweet spoonful (or handful) of our dxe9licieux desserts as much as we enjoyed making them for Merci! Click here for more info on Petit Pot and our gourmet French Create a free website Powered

Table B6: Five Closest Startups and Public Companies to **olomomo.com**

Type	Similarity	Website	Website Text (first 500 characters)
public firm	.568	www.inventurefoods.com	Fridays xab Inventure Foods Innovative + = Inventure Our Brands Boulder Canyon Natural Foods Jamba Rader Farms Fridays Burger King Poore Brothers Tato Skins Braids Pretzels Bobs Texas Style Where To Buy In The News About Inventure In The Community Our History Careers Investor Relations Contact Us Contact Us Product Feedback Donation Requests Home Browse Home Our Brands Fridays Boulder Canyon Natural Foods Jamba Rader Farms Fridays Burger King Poore Brothers Tato Skins Braids Pretzels Bobs Tex
public firm	.548	www.armaninofoods.com	Armanino Foods - Home Home About Us Grocery Foodservice Industrial Recipes News FAQ Contact Us Welcome to Armanino Foods! amilies and businesses have depended on the quality of Armanino products since the , when Guglielmo Armanino provided fresh vegetables and herbs to the growing Italian community in San By the 50s, the family had adopted and a few years later became a forerunner in providing frozen convenience to consumers and the food Armanino Brand Pesto, Pasta and Meatballs have been favo
public firm	.548	www.montanamills.com	Montana Mills Visit Our Bakeries Find a Montana Mills Near You! Supporting Your Communities What s New Monthly Calendar Bread Schedules Montana Mills Kids To Home Page Shop by Price Choose Under $25 Under $50 Over $50 Call for more information (Privacy Statement) This site powered by xa9 Copyright Montana Mills When you visit any Montana Mills Village Bakery, you ll be greeted by one of our friendly crewmembers with an offer for a free slice of our delicious bread (”Slices the size of Monta
public firm	.518	www.pyramidbrew.com	Pyramid - Press Our Brews All About Beer Alehouses Distributors Company Press Contact PRESS Press It can be hard to keep up with all the fun were This page will give you a fighting Here youll find the latest information about our new products, seasonal beer releases, awards and Have at PYRAMID BREWERIES CHANGE OF DIRECTORS August 5, Pyramid Breweries Announces Resignation of Officers and Directors PYRAMID TENDER OFFER CLOSE August 4, Independent Brewers United Announces Completion of Tender Of
public firm	.497	www.webvan.com	Herbs, Spices and Seasonings — Webvan Home — View Cart — My Account — Order Status Search Browse All Categories Baking Supplies Beverages Boxed Meals and Side Dishes Breads Breakfast Foods Coffee, Tea and Cocoa Condiments, Sauces and Spreads Dairy and Eggs Desserts and Pastries Fruits and Vegetables Health and Family Herbs, Spices and Seasonings - Herbs and Spices - Mixed Spices and Seasonings - Pepper and Peppercorns - Salt and Salt Substitutes Home Brewing and Wine Making Household Supplies Meats Pasta and Grains Pet Sup
startup	.496	foodzie.com	Products Tagged with Foodzie: Taste Something Different sign in — help Home Blog Items in your cart Checkout 0 Home Tags All Tags: artisan dessert coffee butter aged chocolate cheese meat creamy gum vanilla vegan mint cookies sweet truffle organic cinnamon orange tea beverage caramel nutmeg raspberry lemon ginger gourmet popcorn snack all natural heirloom sustainable loose tea oats swedish spice granola breakfast cereal bacon gift chile seasonal marshmallow farmstead raw cheddar chipotle gouda
startup	.495	1000corks.com	Corks - Wine Search Corks Also search wine description What s Corks? A powerful but easy to use wine search We search the top wine stores in the United Find any wine, from common to Sort the results by price or Buy wine online, or at your local About Blog Popular Searches: Altocedro Dominus Kosher Cabernet Torre Muga Yquem Sign Up For Our Newsletter Stay up to date with the latest wine releases and Plus, get exclusive discount codes at the best wine Enter your email: Copyright xa9 All Rights
startup	.452	www.shearadiance.com	Shea Radiance, natural and organic shea butter inspired formulations for bath and body Home New Products All Products Specials Search Shop SHOP BY PRODUCT Ultra Rich Body Creams Whipped Shea Body Butters Pure Shea Butter Black Soap - Alata Samina Shea Butter Therapies Aromatherapy Oils Body Scrubs Hair and Scalp Mother and Baby Soaps - Organic SHOP BY SCENT Citrus Floral Fruity Unscented Earthy Spicy COLLECTION Men Mother and Baby Aromatherapy Shea Therapies ACCESSORIES GIFTS Gifts Under $35.00 Gi
startup	.443	santasti.com	SanTxe1sti the palate cleansing beverage HOME — NEWS — PRODUCTS — Txc1STI TIPS — OUR STORY — WHERE TO FIND — ORDER SANTxc1STI — CONTACT US LEARN MORE SITE MAP PRODUCTS FIND SANTxc1STI ORDER SANTxc1STI Txc1STI TIPS WHO IS SANTxc1STI? Quote of the Month The wine industries new friend! An amazing, yet simple way to keep your palate As industry professionals, we drink a lot of wine and SanTxe1sti works to clean the palate in a refreshing way like nothing else does! Allison Dominguez Owner
startup	.439	liquor.com	Grenadine Become a VIP Signup for Discounts And Specials Home Store Recipes Buzz SHOP FOR LIQUOR LIBATION LIBRARY NEWS + MORE Home Grenadine Search Drink Recipes Drink Recipes Recipes By Ingredient Brandy Rum Vodka Whiskey Tequila Gin Recipes By Type Cocktails Cocoa Coffee And Tea Liqueur Punches Shots Most Viewed Drink Recipes Ball Candy Gold Monkey Bora On The Of July Lemonade And Sprite New Drink Recipes Bocci Ball Bogomip Bold Gold Monkey Bomb Bora Bora Born On The Of July Bottlecap Bourbo