Buy, Keep or Sell: A Market for Startups

Mufitcan Atalay

April 2020

1 Introduction

Start-ups are becoming an increasingly larger part of our lives, both as individuals and in the economy. This is true both in terms of the sheer number of start-ups but also in terms their increasing influence in the economy. For example, in 2011 there were less than 20 unicorns in the United States, but in 2018 alone 28 new companies became unicorns for a total of 138 unicorn companies[1]. The behaviour of start-ups in relation to each other, therefore, has become an important idea to explore.

A behaviour of large start-ups that has drawn scrutiny from both regulators and the media is the acquisition practices of large start-ups. For example, since 2003 Google has averaged 11.9 acquisitions a year and 18.625 acquisitions a year since 2010[2]. This scrutiny was not unfounded, as the trend of an increasing number of acquisitions is quite prominent in figure 1, particularly after 2015.

Existing literature on start-up acquisition has mostly revolved around entrepreneurial decision making and the entrepreneur rather than the firm. A good example of this type of literature is Arora et al. 2018 where the timing of exit (selling the start-up) is a strategic choice that determines the funding strategy of an entrepreneur. Firm-level analysis of start-up acquisitions that try to identify patterns and attributes has been largely absent from the literature.

[1] https://pitchbook.com/news/articles/the-us-unicorn-boom-is-reaching-new-heights-in-2018

[2] https://www.wired.co.uk/article/google-acquisitions-data-visualisation-infoporn-waze-youtube-android

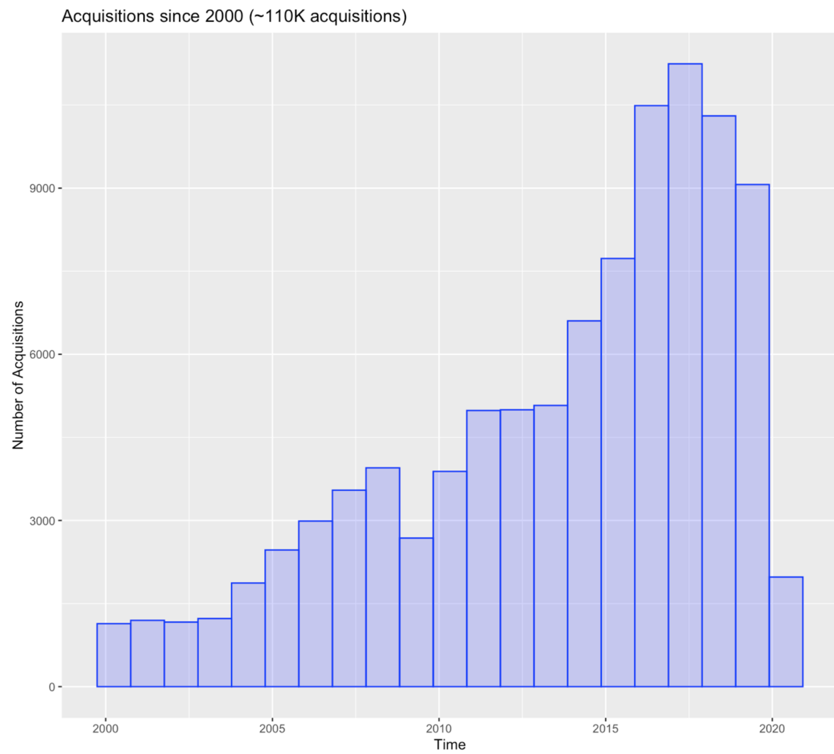

Figure 1: Acquisitions since 2000 from CrunchBase dataset

In this respect a firm-level analysis of acquisitions will have to use a novel data set and to draw a model from other literature. In this case, I will use a data set from crunchbase.com containing 113,762 companies to use a similar model to that of Akcigit, Celik and Greenwood (ACG). The CrunchBase data set labels each firm with multiple industry labels and contains firm-level financial information such as funding amounts and acquisition dates. This will allow us to construct a distance metric between firms and the initial workings of a model that is similar fashion to ACG.

2 Empirical Facts

CrunchBase is a data-as-a-service website where individuals and firms can learn obtain information about start-ups and firms. Their data is anonymously provided and verified by their own personnel. I was able to extract the data set from the website by buying a 1-month subscription to their *Pro* service for 30 dollars. Using the CrunchBase data set we can draw out some empirical facts which will both motivate the investigation and provide evidence for certain assumptions in the model. These empirical facts are listed below.

Empirical Fact 1: *The Number of Acquisition have been increasing*

This fact is clearly established in figure 1 as the number of acquisitions roughly doubles every 5 years after 2000. As outlined in the introduction this empirical fact is important to establish because it justifies an investigation into acquisitions.

Empirical Fact 2: *On Average, Acquiring firms have received a larger amount of funding than Acquiree firms*

Figures 2 and 3 contain the distributions of the amount of funding for acquiree and acquiring firms respectively. The x-axis is on log-scale for the purposes of visualization, but it is still clear that the mean of the two distributions are different even though the shapes of the distributions are very different. This fact is important in the sense that it provides justification for distinguishing between Acquiring and Acquiree firms, particularly in terms of whether a firm's ability to acquire another firm.

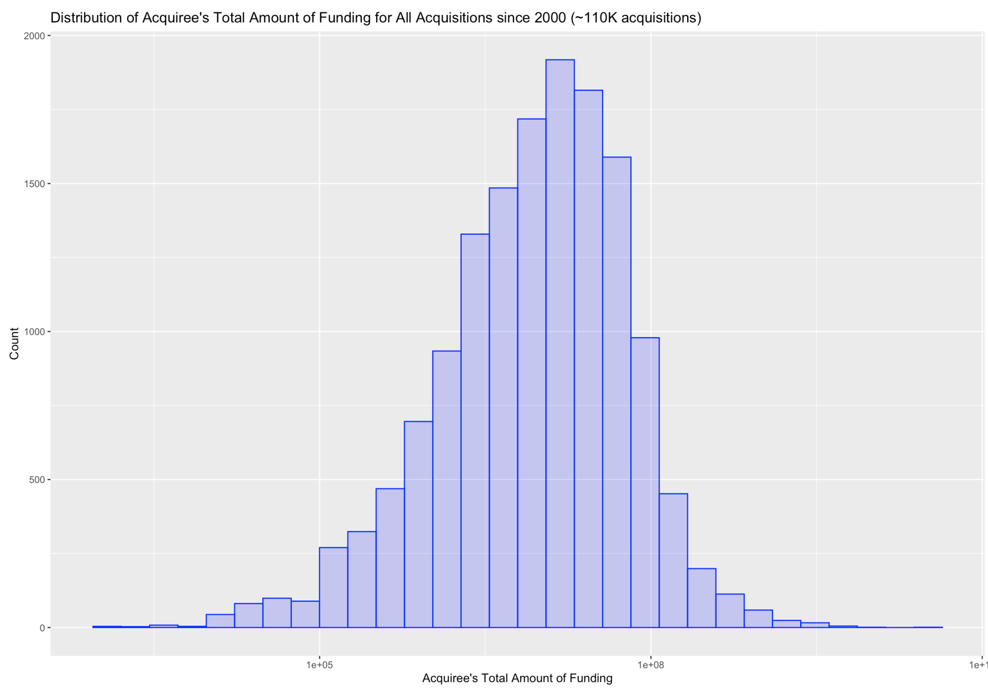

Figure 2: Distribution of Funding for Acquiree

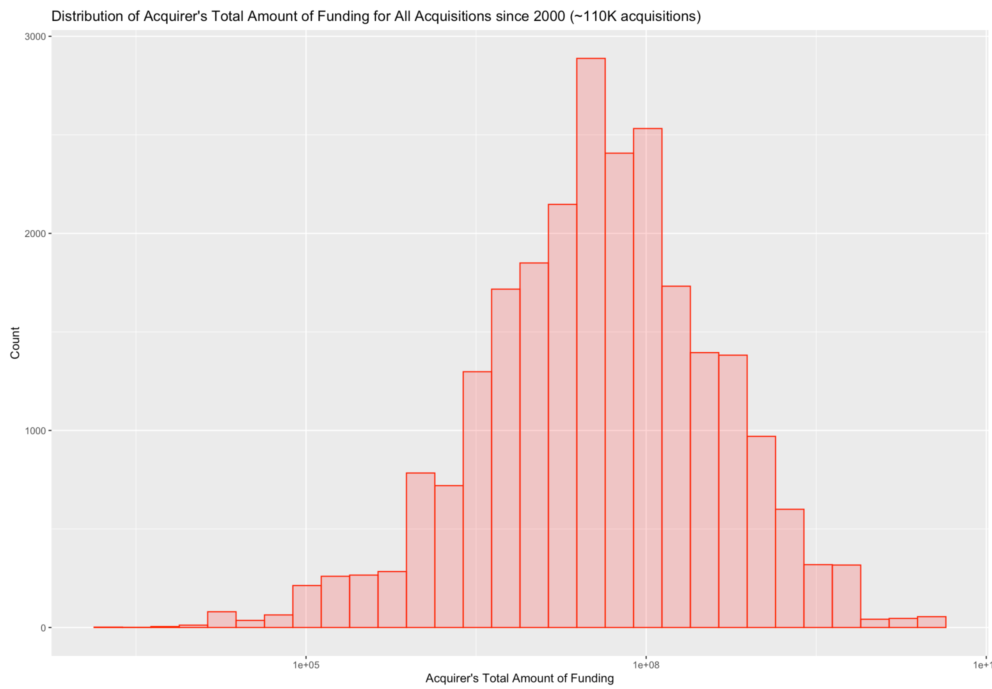

Figure 3: Distribution of Funding for Acquirer

Empirical Fact 3: *Most firms are acquired before first 3 rounds of funding*

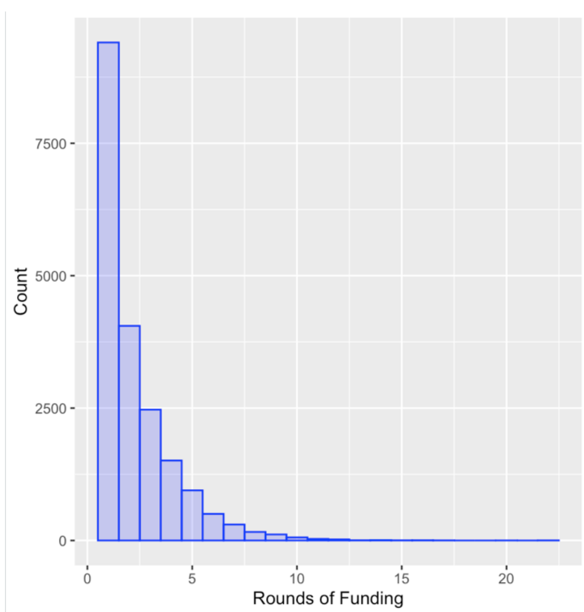

Figure 4: Rounds of Funding Before Acquisition for Acquiree Firms

As we can see from figure 4, firms acquiree firms are often acquired before they reach later stages of funding. In fact most firms are acquired after their first round, and almost 60% of firms are acquired before their first 3 rounds. This is highly important fact for two reasons:

1. This is evidence that **young** firms are being acquired. This is particularly important because it seems to justify the aforementioned scrutiny on acquisitions by large firms and worries that large firms are using acquisitions to shield themselves from possible small competitors and stifling innovation.

2. This is evidence that Venture Capital Firms are not just investors in start-ups, but are also mediators in the market for start-ups, just like patent lawyers in ACG. It seems as though other start-ups find out about new start-ups, or at least find out about the quality of new start-ups from venture capital firms and decide to acquire or not to acquire based on this information. This is also in line with existing entrepreneurial decision making literature where entrepreneurs decide who decide to exit early opt for a short funding strategy[3].

3 Structure of Model

The key idea of the model is to endogenize the acquisition of a firm based on distance like ACG with patents and also valuation, but the valuation of a firm itself will not endogenized. The reason that the model will be able to avoid endogenizing valuation is because of empirical fact 4. Since no existing firm will know the valuation of a new firm (specifically it's post-money valuation) before it receives funding from a venture capital firm, we will be able to just assume that the initial valuation of company is drawn from the distribution $V()$ that will be disciplined by the distribution in Figure 2.

Existing firms will then decide to either acquire the new firm based on the new firm's valuation v_i $V()$ and the distance between itself and the new firm. This process is illustrated in figure 5, where an existing firm f_j will decide not to buy firm f_l because the distance between the two firms is too large, will be unable to buy firm f_i because the new firm have too high of a valuation in comparison to itself and will thus not be able to afford it and will ultimately decide to buy firm f_k because it is sufficiently "close" and sufficiently smaller in valuation in comparison to itself.

[3] Arora et al. 2018

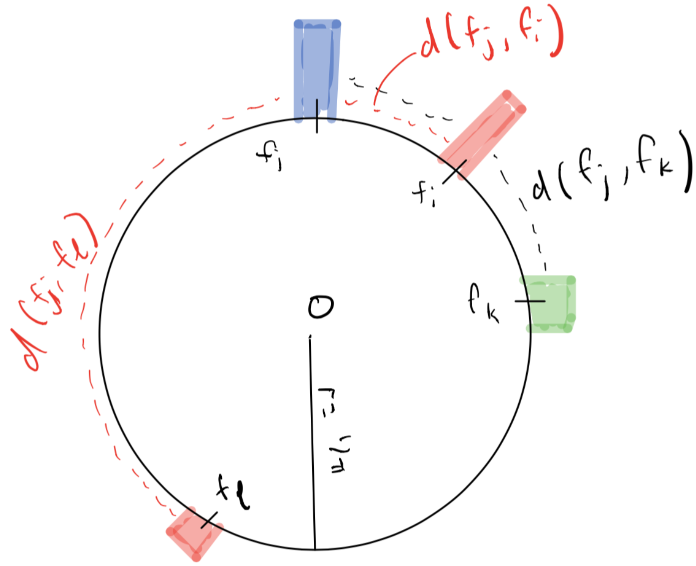

Figure 5: Circle of the Market for Startups

For both existing firms and new firms that are not acquired, they will grow at a higher growth rate λ_h or lower growth rate λ_l and will be faced with a set of decisions:

1. Firstly, they can exit (firm can go bankrupt) at an exogenous rate δ_i that depends on the growth rate in the previous period. The idea here is that firms can be hit with exogenous shocks to their industry or the whole economy and more successful firms will be able less affected from these shocks.

2. Secondly, they can decide to get acquire another existing or new firm based on the process described in figure 5. The next period the valuation of the acquirer firm will become its valuation plus the valuation of the acquiree firm.

3. Thirdly, they can decide to just keep going and grow at either λ_h or λ_l. Plans for the future will explore endogenizing the probability of each type of growth based on valuation or age of firm, although it's possible that this might over complicate the model.

4. Fourthly, they can get acquired and receive its valuation based on the growth in the previous period. There is merit to introducing an acquisition premium which is estimated from the data, which will be explored in the future.

Since this process will occur recursively, the mode lends itself very well to a value function formulation of the problem just like ACG. Below is Figure 6 where the timeline of events described in the previous paragraph is displayed.

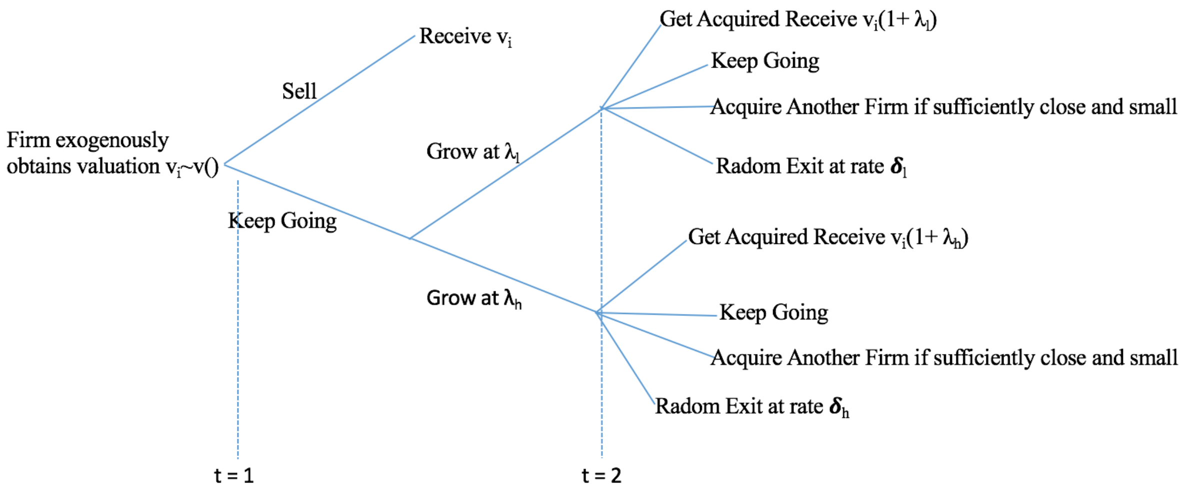

Figure 6: Timeline of Events

4 Distances between Firms

The distance between two Industries is given by:

$$d(I_i, I_j) = 1 - \frac{\#Firms\ with\ Industries\ i\ and\ j\ simultaneously}{\#Firms\ with\ Industries\ i\ or\ j} = 1 - \frac{\#(i \cap j)}{\#(i \cup j)}$$

Since CrunchBase gives on average 4.6 labels to each firm we can construct a distance between two firms by using the distances between the industry labels of the two firms:

$$d(f_i, f_j) = [\frac{1}{||I_{f_i}||}\frac{1}{||I_{f_j}||}\sum_{'I_i \in I_{f_i}}\sum_{'I_j \in I_{f_j}} d(I_i, I_j)]$$

Where I_{f_i} is set of all industry labels for firm f_i. The distance between two firms is now the average distance between the industry labels of the two firms.

From Figure 7 we can see that the distance between acquirer and acquiree firm for the top 100 acquisitions by value is almost always less than 0.2. This is a very good sign because the small and consistent distance between acquiree and acquirer firms shows that firms acquire other firms in closely related industries rather than firms that are "far away". This is evidence that the use of the distance metric in a fashion similar to ACG is reasonable in this context. In Figure 8, we can see that this same pattern holds (and also a rough upper bound of 0.2) for the top 1000 acquisitions as well so we have even stronger evidence that this distance metric is applicable in this context.

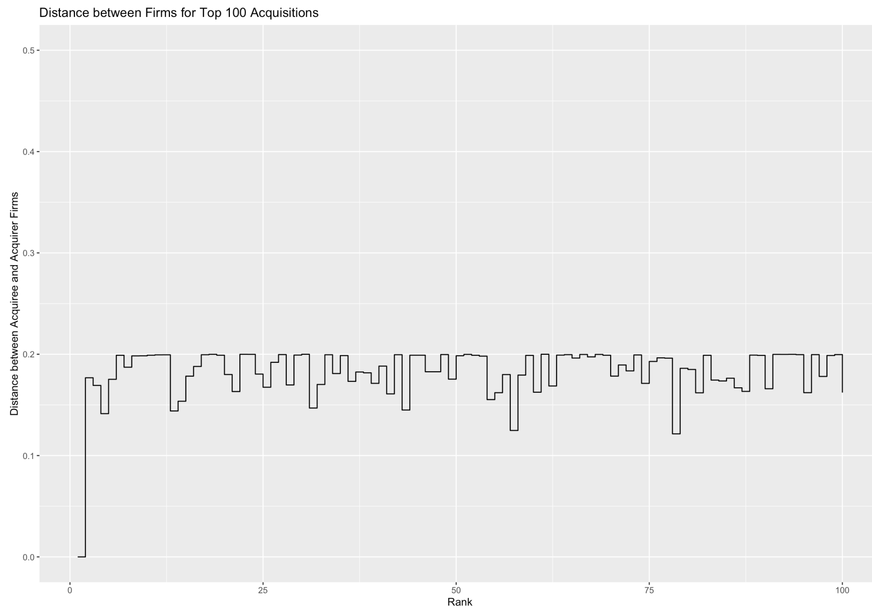

Figure 7: Distance between Firms for Top 100 Acquisitions by Value of Acquisition

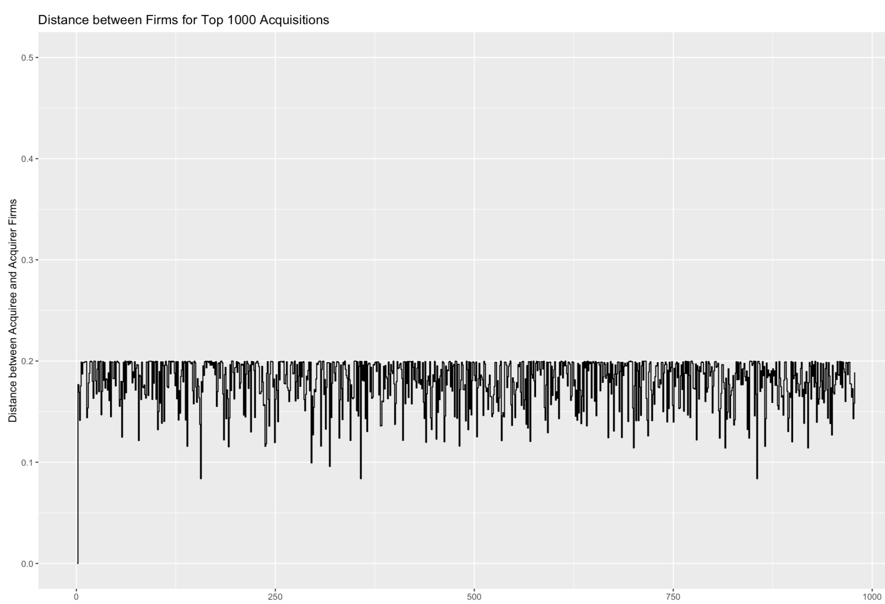

Figure 8: Distance between Firms for Top 1000 Acquisitions by Value of Acquisition

5 Shortcomings of the Data

There are several shortcomings of the data that creates some biases in the figures shown above. Although these biases are less influential in the figures necessary to establish empirical facts, it may affect the distance between firms in a more substantial way.

Firstly, although the data set is quite expansive it does not contain **all** the acquisitions during this time period; however, the data set is probably large enough such that it is as close as possible to a representative sample. There might be a slight upward bias because the data set, most likely, does not contain a representative proportion of acquisitions of smaller sized firms that took place between 2000-2020, but since this affects both acquiree and acquirer firms it should not affect our empirical facts substantially. It's possible that this upward bias does effect the distributions of distances as a whole as smaller firms are going to be less likely to acquire firms a larger distance away from them.

Secondly, the CrunchBase data set does not contain funding information about every single firm that is an acquiree or an acquirer. This is particularly the case for acquirer firms because a non-insignificant portion of the acquirer firms are firms are more traditional and would not be considered as a "start-up" at any point in their lifetime. This means that there are probably very large acquirer firms that are observed as receiving little to no funding and that the difference between the distributions of "funding" (or valuation) acquiree and acquirer firms is probably larger than it appears in fact 2. In this sense this is a positive shortcoming because it gives more validity to the idea that larger firms are buying smaller firms. Another drawback of this shortcoming is that it can mean that regressions and analysis concerning the funding/valuation of a firm and the distance between its acquiree and acquirer cannot be reliably done.

Thirdly, a small portion of acquiree firms (roughly 2000) do not contain information about their number of rounds of funding. It is unclear whether these start-ups received no funding and were acquired by a firm or whether CrunchBase doesn't have this piece of information for this firm. The former of these two options is slightly concerning because it weakens the assumption that other start-ups find out about new start-ups through venture capital firms. On the other hand the fact that this is a small amount relative to whole size of the data set is encouraging for the strength of this assumption.

Fourthly, new labels for industries are constantly being created in the CrunchBase data set and this upwardly biases the calculations between firms. For example, there are no "Marketing Analytics" firms acquired before 2012, however there are "Advertising" firms being acquired and acquiring all throughout the time span of the data set. The distance between these two industries is 0.38, but this is also the distance between "Advertising" and "Internet" labels. In reality almost all "Marketing Analytics" firms are labeled with "Advertising", but the number of "Advertising" labeled firms is so much larger and the label is so much older, it can make industries which are actually a lot "closer" have a larger distance and labels that have existed since 2000 to appear "closer".

6 Future

There are two main elements of this proposal that require serious developments in the future. The first element is to formalize and solve the model. This will most likely require the formulation of a value function similar to ACG without the matching element. Potentially, It could be possible to endogenize the probability of an arrival of λ_h vs. λ_l through endogenous labour choice by firms; however, this could really mess-up the timing of certain decision by firms. In addition the endogenization of firm growth rates is not the goal of this proposal.

The second element is to improve or fix the shortcomings of the data. The main way to fix the first 3 shortcomings would be to use a larger more comprehensive data set than CrunchBase. This might be slightly difficult because a significant portion of firms involved in these acquisitions are private firms. In this sense CrunchBase's data might be as good as it gets. The solution to the 4th shortcoming probably requires some communication with CrunchBase because it is not possible to first occurrence of each label because new labels have been retroactively assigned to firms (for example Google has the label "Entertainment" even in its first acquisition in 2001 when it had not yet acquired YouTube).

Some interesting extensions could include:

- Changing the distribution of acquiree and acquirer firms as counterfactuals
- Concentrations of new firms in a specific industry as counterfactuals
- Incorporating micro-data about individual venture capital firm investments and creating a network of firms connected to a venture capital firm. This could allow for the construction of a similar distance metric between different venture capital firms and a "learning curve" for between start-ups. For example if a new firm and an existing firm have the same investor, they may find out about or finalize their valuation of each other immediately. If a new firm and old firm are 2 investor degrees away from each other (investor in new firm also invests in another firm that has the same investor as an older firm) then it may take longer or require more rounds of funding for the older firm to find out or finalize their valuation of the new firm.
- Compare and contrast different distance metrics, possibly some machine learning algorithms like K-means clustering or random forest.

7 References

Arora, Ashish, Andrea Fosfuri, and Thomas Roende. "Waiting for the Payday? The Market for Startups and the Timing of Entrepreneurial Exit," 2018. https://doi.org/10.3386/w24350.

Akcigit, Ufuk, Murat Alp Celik, and Jeremy Greenwood. “Buy, Keep, or Sell: Economic Growth and the Market for Ideas.” Econometrica 84, no. 3 (2016): 943–84. https://doi.org/10.3982/ecta12144.