Patterns of Successful Founding Team Composition and Funding Outcomes

Francesco Ferrati[1], Phillip H. Kim[2], Moreno Muffatto[1]

[1]University of Padova, School of Entrepreneurship (SCENT), Department of Industrial Engineering
[2]Babson College, Arthur M. Blank Center for Entrepreneurship

francesco.ferrati@unipd.it (corresponding author)
pkim1@babson.edu
moreno.muffatto@unipd.it

Abstract. When it comes to assessing a startup's chance of success, equity investors apply a specific set of criteria to minimize risk. In their decision-making process, most venture capitalists (VCs) agree with giving priority to the team composition, hence the popular saying: "*Always consider investing in a grade-A team with a grade-B idea. Never invest in a grade-B team with a grade-A idea*." In this paper, we explore the profile of technology-based startup teams that are most likely to secure a Series-A funding round from VCs. From a methodological point of view, we applied a strongly quantitative approach, integrating several data mining techniques according to a multidisciplinary perspective, between data science and entrepreneurship. As for the company information, we used Crunchbase as our primary source, considering a set of U.S.-based startups founded from 2000 to 2017. For each venture we algorithmically integrated team-related information from the founders' public LinkedIn profiles. Overall, we analysed more than 2,100 teams, involving a total of about 4,600 founders. Each founders' experience was analysed by considering their professional background. Overall, more than 29,000 work experiences have been taken into consideration. Statistical analysis was carried out on both individual founders and their team organization. Both founders and teams were evaluated in terms of heterogeneity of prior experience and similarity of co-founder profiles using the Gini coefficient and Jaccard index, respectively. Statistics are expressed according to the companies' sector and their fundraising profile. In fact, the different sectors are mapped on a 4-quadrant chart to identify different combinations between founders' profiles (specialists VS generalists) and teams characteristics (combining co-founder with similar or diverse background). Results reveal the impact of team similarity and variety in terms of prior working experience. The findings provide valuable insights for scholars dealing with tech-driven startups teams, aspiring entrepreneurs looking for co-founders and for VCs seeking to invest in promising startups.

Keywords: startup, team, venture capital, data mining, linkedin

1. Introduction

In the highly competitive and rapidly evolving landscape of technology startups, the role of the founding team has emerged as a critical determinant of success (Muñoz-Bullon et al., 2015).

Over time, scholars in Entrepreneurship have explored different perspectives for analysing the characteristics of a founding team. The main areas of investigation include human capital, social capital, psychological traits, and team variety. Human capital theory (Becker, 1964; Unger et al., 2011) highlights the importance of founders' experience, education, and skills as signals for a team's potential for future success. Social capital theory (Aldrich and Zimmer, 1986; Shane and Cable, 2002) emphasizes the role of networks and relationships in facilitating access to resources and opportunities. More recent research has investigated the role of the founding team's psychological traits (Koellinger et al., 2007; Hmieleski and Baron, 2009), such as risk-taking propensity, tolerance for ambiguity, and self-efficacy. Finally, increasing emphasis is being given to the team diversity (Ruef et al., 2003; Bouncken et al., 2016) in terms of demographics, cognitive styles, and values, in determining the success of startup ventures. In this paper we focus on both human capital and team variety, considered from the work experience perspective (Hashai and Zahra, 2022). Specifically, human capital is analysed starting from the career path of each individual founder. On the other hand, variety within the team is considered in terms of heterogeneity and complementarity of the co-founders' professional background.

When discussing about human capital and variety of founders' profiles, two types of career paths can be outlined: so-called specialists and generalists (Custódio et al., 2013; Kaiser and Müller, 2015, Xiao et al., 2020). Generalists have interdisciplinary backgrounds leveraging a broad range of skills and knowledge across multiple domains. Thanks to their cross-sectional expertise, generalists can quickly adapt to new situations and challenges, allowing startups to pivot more easily and leverage new opportunities; they can facilitate communication between different departments, promoting a collaborative work environment and can identify connections between seemingly unrelated fields, leading to innovative ideas and solutions. However, as they act as "Jack of all trades, master of none", generalists may struggle to excel in highly specialized or technical roles that demand focused expertise. On the other hand, specialists own deep knowledge and expertise in a specific domain. They have spent considerable time and effort mastering a particular field, and their skills are often crucial in addressing complex, domain-specific challenges. Thanks to their vertical expertise, specialists are often more efficient in their specific areas, as their deep expertise enables them to work quickly and accurately. However, being focused on a single topic only, specialists may struggle to adapt to new domains or to connect the dots together. In the context of technology-based startups, the choice between generalists and specialists in a founding team can significantly impact the venture's level of success, innovation and adaptability.

One of the first specific situation in which a team is evaluated in depth is the fundraising process. In fact, when it comes to assessing a startup's chance of success, equity investors apply a specific set of criteria to minimize risk (Ferrati and Muffatto, 2021). Specifically, most venture capitalists (VCs) agree with giving priority to the team composition in their decision-making process (Franke et al., 2006; Franke et al., 2008). As VCs highly prioritize the quality of the team as a go/no-go factor, in this paper we explore possible relationships between the characteristics of the founding team and its ability to raise a series-A funding round.

However, the rationale applied by VCs in evaluating the efficacy of a founding team also depend on the specific industry in which the startup operates, due to the specific challenges that characterize each sector. For this reason, in our analysis we also considered each team according to their sectors.

This study thus aims to address the following research question: "*Considering the industry in which a startup operates, how can the human capital of individual founders and the variety within the founding team affect the ability to raise a series-A funding round?*"

From a methodological perspective, we addressed the research by applying a highly quantitative approach. In fact, in both the data collection and the data analysis phases, we applied advanced data mining techniques. As for the ventures' information, we used Crunchbase as our primary data source (Ferrati and Muffatto, 2020). Also, in order to analyse each founder's professional background in detail, for each company we algorithmically integrated team-related information from the founders' public LinkedIn profiles (Gloor et al., 2011; Banerji and Reimer, 2019). Indeed, this paper aims to contribute to the literature by exploring the potential of data science to study team-related topics using complex data sets.

The structure of the paper is as follows. First, we describe the methodological process, from data collection, to professional experiences coding and classification, to the definition of variables and indices. Next, we present the analysis results, highlighting the values of two considered indices (Gini coefficient and Jaccard index) calculated for founders and teams respectively in different sectors. Finally, we discuss the results and present future research opportunities.

2. Methodology

The entire research activity was carried out by applying an advanced data science approach. Data science is an interdisciplinary field that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract valuable insights and knowledge from complex data sets. It involves the entire data lifecycle, including data collection, cleaning, analysis, visualization, and communication of results. In the following sections we describe in detail all the specific phases for the research performed. i.e., data collection, work experience classification, variable definition, index analysis. All operations were performed using the Python programming language.

2.1 Data collection

Since our research approach is based on the analysis of large amounts of data, the choice of the most suitable datasets to use highly influences the quality of the results. Considering the scope of the research, data were obtained from two complementary data sources i.e., Crunchbase and LinkedIn. We used Crunchbase as our main reference for companies. Companies' information was then enhanced by integrating each founder's LinkedIn profile.

Crunchbase is one of the most relevant databases collecting information on technology-driven companies. The data considered was retrieved on February 5, 2021. Overall, the Crunchbase database is organized into seventeen individual .csv files and the following five datasets were considered in this study: *Organizations*, *People*, *Jobs*, *Investments* and *Funding rounds*. Specifically, the *Job* dataset was used to identify all job positions with a "founder" or "co-founder" title. The founders' job positions were then linked to the founders' information in Crunchbase (dataset *People*) and the respective founded companies (dataset *Organizations*). For both companies and persons, Crunchbase provides a variable with the URL of the respective LinkedIn profile. At this stage, both companies that did not report a valid company LinkedIn URL and companies that did not report all founders LinkedIn URL were discarded. An intermediate sample of 83,347 companies was so obtained. Once the set of companies with a complete set of LinkedIn URLs was obtained, their profiles were integrated with their fundraising information (datasets *Investments* and *Funding rounds*). For each company we tracked the entire funding history, in terms of the number of funding rounds, raised amount (in total and in the eventual series-A), and timing (date of the eventual series-A or of the last funding round). At this point we obtained an intermediate sample of 61,172 companies. Companies were then additionally filtered by considering only U.S.-based ventures (26,430 remaining companies). Since the research aims to analyse the dynamics within founders' teams, only companies with number of founders between 2 and 4 were included i.e., companies with solo-founders were excluded (8,562 remaining companies). Finally, an additional filter was applied by considering the year of foundation of the companies. Specifically, we considered companies founded between 2000 and 2017. The year 2017 was selected by starting from the year 2020 (note that we extracted Crunchbase data in 2021, so we applied a principle of prudence) and subtracting 3 years. The value of 3 years was calculated on the sample as the average number of years to collect a series-A round. In this way, we were sure that companies with no series-A had already passed the average number of years (i.e., 3 years) to collect a series-A. At this point we obtained an intermediate sample of 6,826 companies. The data provided by Crunchbase are mainly related to the companies' registry (location, year of foundation, industry) and their fundraising track record. On the other hand, founders-related information is not sufficiently detailed to allow in-depth analysis of professional backgrounds.

For this reason, we used the list of founders obtained from Crunchbase as the starting point for the extraction of the founders' public LinkedIn profiles. For this purpose, a custom algorithm was implemented to automatically search and collect the founders' public profiles. It is important to note at this stage that only publicly available information was used. It is also crucial to point out that the use of only public data actually led to a significant reduction in the final sample size. Indeed, several cases can occur at the time of extraction that actually result in a discard of at least one founder's profile and consequently of the related company. Examples of cases that lead to the discarding of a founder's profile are: public LinkedIn page not accessible, accessible but empty person public profile (no information provided), incomplete sections (e.g., education section is available, but there is no information about the professional experience). Due to these cases, the final sample used in the study was therefore much smaller than the one initially obtained from Crunchbase. The final sample consists of 2,109 companies (i.e., founding teams), 4,673 founders' profiles and 29,694 founders' prior work experience.

Finally, in order to consider the timing of the fundraising process, we also used the investment date provided by Crunchbase to classify companies according to three possible classes of outcomes. Table 1 summarizes the number of teams (i.e., companies) and founders related to each outcome.

Table 1: Number of teams (i.e., companies) and founders related to each fundraising outcome

Class	Description	N teams	% teams	N founders	% founders
Outcome 0	Companies with no series-a	1,194	56.61%	2,558	54.74%
Outcome 1	Companies that raised a series-a within 3 years from foundation	461	21.86%	1,080	23.11%
Outcome 2	Companies that raised a series-a after 3 years from foundation	454	21.53%	1,035	22.15%
Total		**2,109**	**100.00%**	**4,673**	**100.00%**

2.2 Work experience classification

Once the complete dataset of companies with their founders and related work experience was created, an intensive reclassification activity was required to analyse the professional backgrounds. Within the LinkedIn profile, all individual work experiences have the same format consisting of: the title of the experience, the name of the organization, the duration (start date - end date - duration), the geographic location, and a brief description of the activity (please, note that not all information is always fully provided). The key information (also mandatory) for understanding the specific activity is the so-called job title. Due to the way this variable is designed, its analysis is particularly challenging. In fact, this field is designed as a "free text" data type and actually allows the user to write what they consider most informative for the job identification. Thus, considering the high variability of this variable, a reclassification of the possible values toward a manageable but significant number of experience areas was first necessary in order to perform a statistical analysis.

In order to classify the variety of existing occupations in a consistent way, several governments have developed different occupational taxonomies (e.g., SOC for US occupation and NOC for Canadian occupations). An advanced job classification system is provided by the O*NET Program, the U.S. nation's primary source of occupational information (National Center for O*NET Development, 2023). The O*NET database, contains standardized and occupation-specific descriptors on almost 1,000 occupations covering the entire U.S. economy. The considered taxonomy is based on the 2018 SOC system (i.e., Standard Occupational Classification, by U.S. Bureau of Labor Statistics) but also provides several grouping schemes that combine different occupations together based on multiple aggregation criteria. Specifically, the 16 "career clusters" groups together occupations within the same field that require similar expertise. As in this research we wanted to explore the impact of the heterogeneous/homogeneous profiles of the founders on the investments collected, O*NET's "career clusters" are a valuable rationale for experience classification.

The LinkedIn experience reclassification process has therefore resulted in categorizing each LinkedIn "job title" into one of the 16 possible "career clusters". It was therefore necessary to create an algorithm to let the two systems communicate with each other. The reclassification system we developed propose an original search strategy for specific keywords within the extended job title stated on LinkedIn. The system was designed from an analysis of job titles' format which generally consists of 3 sections: type of contract, responsibility, and area. As in Figure 1, in the case of "*full-time lead software developer*", "full-time" represents the type of contract, "lead" describes the responsibility, and "software developer" is the occupation area.

Figure 1: Breakdown strategy for job titles analysis

Since for the purposes of our research, the most important element is the "area", a search strategy focused on optimizing the retrieval of this information was applied. Specifically, our algorithm was structured in 3 sequential phases: (1) searching for any "contract type" information, (2) searching for any "responsibility" information, and finally (3) searching for any "area" information. Where the first 2 steps were successful (i.e., the algorithm found information on contract type and/or responsibility), the part of the text referring to them was dropped from the title string. The third step (i.e., "area" identification) was then applied only to a portion of text that should most likely contain information about the area itself. This strategy has made it possible to eliminate much of the informational noise present in the job titles on LinkedIn. In each of the three steps of the algorithm, to search for each piece of information, the possible classes of each part were first identified and then a comprehensive list of keywords related to each class was created. In total, more than 850 unique keywords have been considered. The algorithm searched for all the keywords associated with each class and, where found, assigned the corresponding class to the portion of text. Table 2 shows the different possible classes that can be associated with the three types of information.

Table 2: Possible classes that can be associated with each part of a job title

Job title parts	Possible classes
Type of contract **7 possible classes**	*full time · part time · contract · independent contractor · temporary · on call · volunteer*
Responsibility **13 possible classes**	*founder · strategy · executive · senior management · middle management · specialist · associate · entry level · student · researcher · advisor · trainer · other*
Area i.e., O*NET "career clusters" **16 possible classes**	*agriculture, food & natural resources · architecture & construction · arts, audio/video technology & communications · business, management & administration · education & training · finance · government & public administration · health science · hospitality & tourism · human services · information technology · law, public, safety, corrections & security · manufacturing · marketing · science, technology, engineering & mathematics · transportation, distribution & logistics*

Figure 2 shows how the job titles of the 29,694 experiences extracted from LinkedIn were uniquely associated with one of the 16 O*NET "career clusters". As can be seen, more than 21% of all the founders' experiences can be associated with the area of "business, management & administration", followed by "information technology" (20%), "science, technology, engineering & mathematics" (15%), "marketing" (14%), "finance" (8%) and "arts, audio/video technology & communications" (6%). All other "career clusters" together collect less than 10% of the experiences while in 6% of the cases the algorithm didn't find any keywords related to any cluster, confirming the vast variability of the job titles on LinkedIn (result "unknown"). Note that in the latter cases of uncertainty, we prudently preferred to associate an "unknown" label rather than enforcing the association of an incorrect cluster.

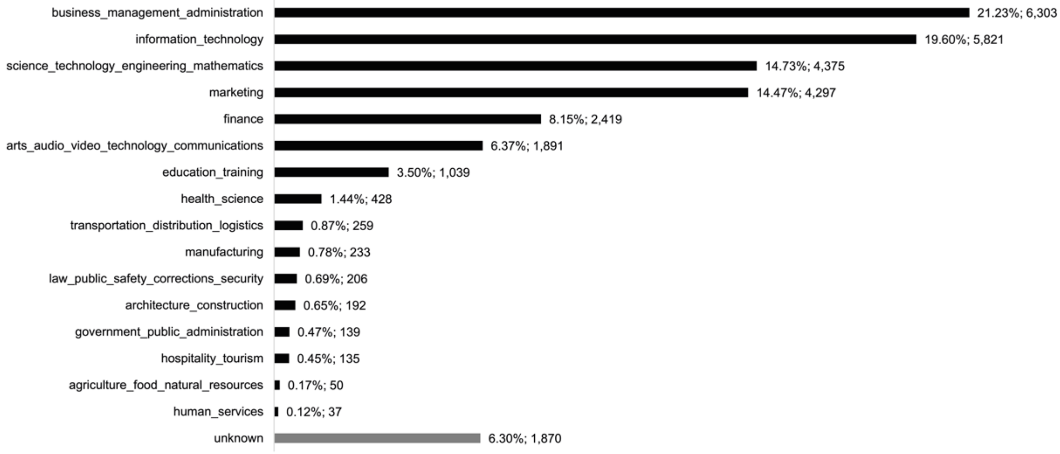

Figure 2: Percentage and number of job titles on LinkedIn related to each O*NET "career clusters"

2.3 Variable definition

The experiences classification process was the starting point for defining the set of key variables to analyse. Specifically, from each founder's LinkedIn experience, the classification algorithm allowed us to identify the number of experiences each founder had in each of the 16 "career clusters". Since we also knew the duration of each experience, the number of months of cumulative experience in each of the 16 "career clusters" was also calculated for each founder. At this stage it is important to emphasize that both any experience of the founders after the foundation of the considered company and the experience itself in the founded company were excluded from the computation. In fact, our research wanted to focus on the analysis of the work background prior to the startup's foundation. Finally, in order to consider the level of heterogeneity and homogeneity of each founder's areas of experience, for each of the 16 "career clusters", the relative duration of experiences was calculated as a ratio of the total duration of the professional career. In this way, each career cluster was associated with the percentage of time the founder spent in that area. Moreover, to allow analysis by sector, the 744 "categories" provided by Crunchbase were also considered. Note that due to the high cross-sectorial activities of tech-driven startups, multiple categories may be associated with the same company. In order to improve the classification system, the 744 existing categories were reclassified into 57 technology-sectors and founders' and teams' profiles were then analysed according to them. Finally, in order to consider the timing of the fundraising process, we also considered the three possible fundraising outcomes as represented in Table 1.

2.4 Index analysis

In this research we wanted to analyse the profiles of the founders and teams in terms of homogeneity or heterogeneity. For this purpose, we decided to summarize the profile of each founder and team through the use of two statistical measures. Specifically, for each founder we computed the value of the Gini coefficient associated with their profile, and for each team we determined the value of the Jaccard index.

The Gini coefficient is a statistical measure traditionally used to quantify income or wealth inequality within a specific population (Equation 1). More generally, the Gini coefficient can be used to measure the heterogeneity of a distribution. When normalized (Equation 2), it ranges from 0 to 1, where 0 represents perfect homogeneity (i.e., data are equally distributed over all k modes, which have equal relative frequency) and 1 represents perfect heterogeneity (i.e., only one of the k possible modes is observed).

Equation 1: Gini coefficient

$$G = \sum_{i=1}^{k} f_i^2 \qquad 0 \leq G \leq \frac{k-1}{k}$$

Equation 2: Normalized Gini coefficient

$$G_N = G\frac{k}{k-1} \qquad 0 \leq G_N \leq 1$$

In our case, we calculated the normalized Gini coefficient for each founder by considering the relative duration of their experiences in each one of the 16 possible "career clusters" (k = 16) expressed as a percentage of the total career duration. In this way, a high G_N value (i.e., closer to 1) models a founder who has gained experience in only one or a few occupational areas (high heterogeneity) and thus has a more specialist profile. On the other hand, a low G_N value (i.e., closer to 0) models a founder who has gained experience in many different occupational areas (high homogeneity) and thus has a more generalist profile.

As for the analysis of the complementarity of the founders within the same team, the Jaccard index was considered instead. The Jaccard index is a statistical measure used to quantify the similarity between two sets (Equation 3). It is calculated by dividing the number of elements in the intersection of the two sets by the number of elements in the union of the sets. It ranges from 0 to 1, where 0 indicates no similarity (no shared elements) and 1 represents complete similarity (identical sets).

Equation 3: Jaccard Index

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

In our case, we calculated the Jaccard index for each team by considering the similarity of co-founders' career, that is, the number of "career clusters" in common among two or more co-founders (each co-founder is considered as a set). In this way, a low *J* value (i.e., closer to 0) models a team whose co-founders have very different professional backgrounds (low similarity) and thus can create synergies across multiple areas. On the other hand, a high *J* value (i.e., closer to 1) models a team whose co-founders have worked in many areas in common (high similarity) and thus can further strengthen the profile they already have individually.

3. Results

In our data analysis, we wanted to explore the relationship between the characteristics of both individual founders and teams as a whole as a function of the company's industry and the outcome of the fundraising process. For this purpose, we carried out a statistic analysis of both the normalized Gini coefficient and the Jaccard index for each of the 57 categories and each of the 3 possible fundraising outcomes.

In appendix, Table A 1 describes in detail the mean of each of the two indices for each sector and for the 3 different outcomes. It should be noted that since multiple categories are usually associated with each company on Crunchbase, the same founders and teams may contribute to the mean of multiple categories. The colour gradient on each column allows to immediately identify the sectors with higher (red) and lower (green) values of each index. Please note that both indexes have minimum value of zero and maximum value of 1. Considering outcome 1 (i.e., companies that raised a series-a within 3 years from foundation), it is interesting to evaluate the sectors that require higher or lower values both in terms of the specialization of individual founders and in terms of the variety of members within the team. For example, the outcome 1 case of the "gaming" sector has a lower G_N value than the other two possible outcomes (equal to 0.36, i.e., more generalist profile of individual founders) and also a lower J value (equal to 0.23, i.e., greater variety of experience among the founders). On the other hand, always considering the outcome 1 scenario, the "life science and biotechnology" sector requires individual founders to have a rather specialized profile (G_N equals 0.45) and similarity among the profiles of the founders (J equals 0.48).

To explore possible insights into the data, considering each of the 3 possible fundraising outcomes, the 57 categories were plotted within a 4-quadrant scatter plot as in Figure 3. The x-axis shows the mean of G_N and the y-axis shows the mean of *J*. Low values on the x-axis represent more generalist founders (having more homogeneous work experiences), while high values on the x-axis represent more specialist founders (having more heterogeneous work experiences). On the other hand, low values on the y-axis represent teams made up of founders with more diverse profiles while high values on the y-axis represent teams composed of founders with more similar profiles. Each point represents a specific sector (the numerical labels correspond to the different sectors as in Table A 1). The dashed lines represent the average value of G_N and *J* calculated over all categories. The 4 graphs in Figure 3 represent respectively (a) all three outcomes with different colours (Outcome 0, 1 and 2), (b) only outcome 0, (c) only outcome 1 and (d) only outcome 2.

It can be seen from the figure that different sectors have different positions in the 4 quadrants. This position also changes according to the fundraising outcome considered. For example, considering outcome 1 (Figure 3.c i.e., companies that raised a series-A within 3 years of foundation), it can be noted that the "transportation" sector (label 53) is located in the slightly-upper right quadrant i.e., more specialist founders and more similar team members. In contrast, the "design" category (label 17) is located in the slightly-lower left quadrant i.e., more generalist founders and teams consisting of founders with more diverse backgrounds. When considered for each sector, this analysis provides a useful benchmark for making an initial analysis of the most characteristic team configurations in different sectors and for different fundraising profiles. However, note that the maximum values of both the normalized Gini coefficient and the Jaccard index are not very high (maximum value 0.61 and 0.58, respectively) suggesting the non-extreme specialization of individual founders and the synergistic complementarity of profiles within the team.

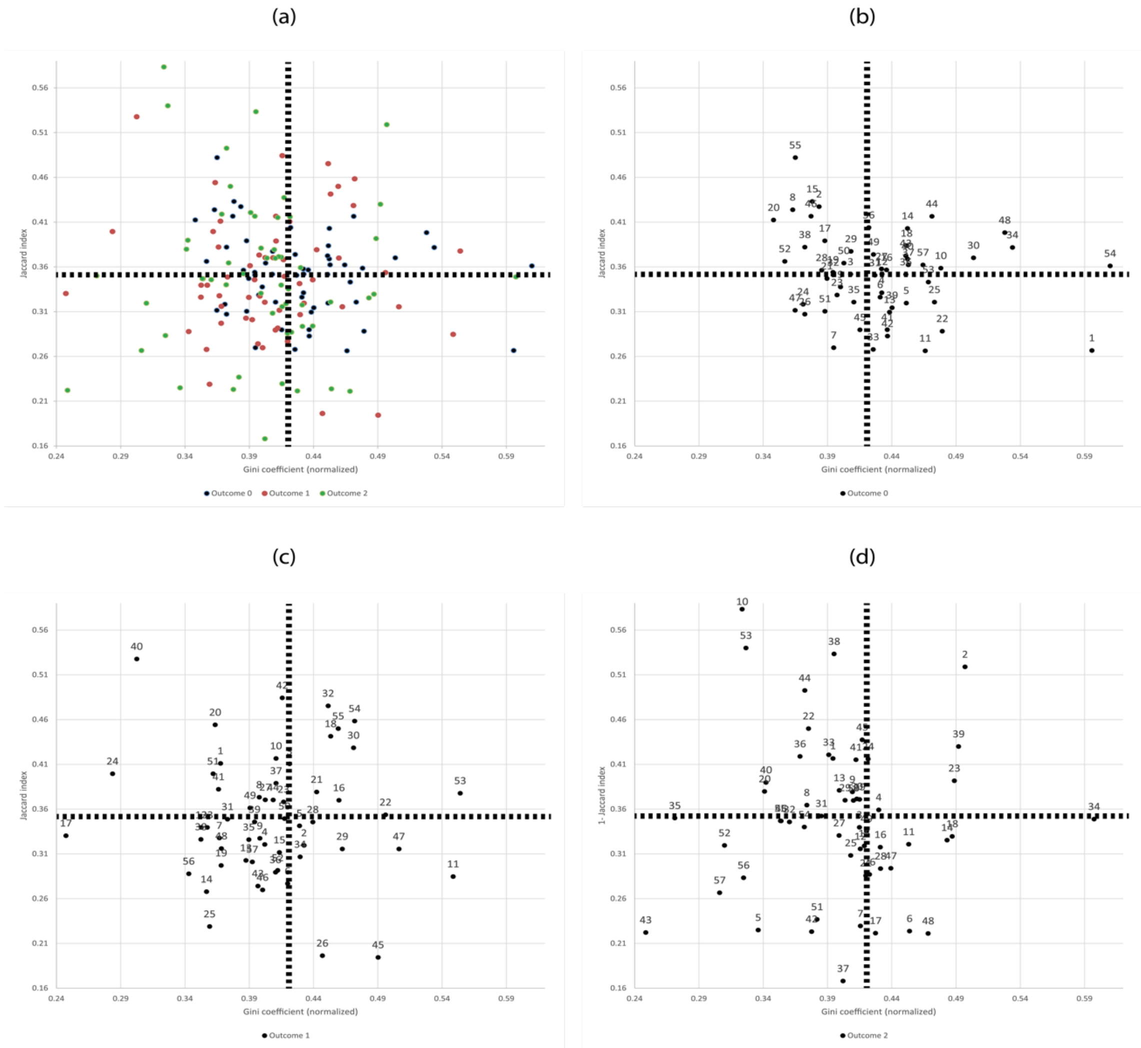

Figure 3: Analysis of Gini coefficient and Jaccard index for both founders and teams of companies in different sectors and with different fundraising profile

4. Conclusion and opportunities for future research

In this research, we proposed an original method to quantitatively analyse the founding team of technology companies. Starting with the identification of target companies on Crunchbase, we reconstructed the professional background of the founders through an analysis of their public LinkedIn profiles. To perform a statistical analysis, the extremely large variety of job titles declared on LinkedIn were mapped back to the 16 possible “carer clusters” of the O*NET program, by using an original classification strategy. The profiles of both founders and teams were then analysed by repurposing the Gini coefficient and the Jaccard index. The results were then plotted considering the mean values of the two indices for different sectors and fundraising profiles.

For the purpose of future research, it is useful to report some current limitations. For example, in the present research we considered only the founders’ professional background, thus not including either the education aspect or the personality aspects of the founders. While education-related information can also be gathered from public profiles on LinkedIn, a large-scale analysis of character and psychological aspects requires a different approach and is certainly very challenging. Another limitation relates to the category classification provided by Crunchbase. Since multiple categories can be associated with each company, the same founders and teams can contribute to the statistics of multiple categories (reflecting, however, the inter-industry aspect of tech-driven startups).

Finally, consider possible future developments, it would be interesting to perform an inferential statistics analysis in order to investigate possible causal relationships between the characteristics of both founders and teams and the companies’ fundraising profile.

Applying a data science approach, the results of our study can provide insight to both scholars on topics related to founding teams, aspiring entrepreneurs in search of a co-founder, and equity investors in their decision-making process.

Acknowledgment

This research was made possible thanks to the support of Crunchbase Inc.

http://www.crunchbase.com

References

Aldrich, H. E., & Zimmer, C. (1986). Entrepreneurship through social networks. In D. Sexton & R. Smilor (Eds.), The art and science of entrepreneurship (pp. 3-23). Cambridge, MA: Ballinger.

Banerji, D., & Reimer, T. (2019). Startup founders and their LinkedIn connections: Are well-connected entrepreneurs more successful?. Computers in Human Behavior, 90, 46-52.

Becker, G. S. (1964). Human capital: A theoretical and empirical analysis, with special reference to education. Chicago: University of Chicago Press.

Bouncken, R. B., Brem, A., & Kraus, S. (2016). Multi-cultural teams as sources for creativity and innovation: The role of cultural diversity on team performance. International Journal of Innovation Management, 20(1), 1650005.

Custódio, C., Ferreira, M. A., & Matos, P. (2013). Generalists versus specialists: Lifetime work experience and chief executive officer pay. Journal of Financial Economics, 108(2), 471-492.

Ferrati, F. & Muffatto, M. (2020). Using Crunchbase for research in Entrepreneurship: data content and structure. In 19th European Conference on Research Methodology for Business and Management Studies: ECRM 2020 (pp 342-351). Academic Conferences and publishing limited, Aveiro.

Ferrati, F., & Muffatto, M. (2021). Reviewing equity investors’ funding criteria: A comprehensive classification and research agenda. Venture Capital, 23(2), 157-178.

Franke, N., Gruber, M., Harhoff, D., & Henkel, J. (2006). What you are is what you like—similarity biases in venture capitalists' evaluations of start-up teams. Journal of Business Venturing, 21(6), 802-826.

Franke, N., Gruber, M., Harhoff, D., & Henkel, J. (2008). Venture capitalists’ evaluations of start–up teams: Trade–offs, knock–out criteria, and the impact of VC experience. Entrepreneurship Theory and Practice, 32(3), 459-483.

Gloor, P. A., Dorsaz, P., & Fuehres, H. (2011). Analyzing success of startup entrepreneurs by measuring their social network distance to a business networking hub. In Proceedings 3rd international conference on collaborative innovation networks coins, Basel. Sept (pp. 8-10).

Hashai, N., & Zahra, S. (2022). Founder team prior work experience: An asset or a liability for startup growth?. Strategic Entrepreneurship Journal, 16(1), 155-184.

Hmieleski, K. M., & Baron, R. A. (2009). Entrepreneurs' optimism and new venture performance: A social cognitive perspective. Academy of Management Journal, 52(3), 473-488.

Kaiser, U., & Müller, B. (2015). Skill heterogeneity in startups and its development over time. Small Business Economics, 45, 787-804.
Koellinger, P., Minniti, M., & Schade, C. (2007). I think I can, I think I can: Overconfidence and entrepreneurial behavior. Journal of Economic Psychology, 28(4), 502-527.
Muñoz-Bullon, F., Sanchez-Bueno, M. J., & Vos-Saz, A. (2015). Startup team contributions and new firm creation: The role of founding team experience. Entrepreneurship & Regional Development, 27(1-2), 80-105.
National Center for O*NET Development. About O*NET. O*NET Resource Center. Retrieved May 5, 2023, from https://www.onetcenter.org/overview.html
Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The structure of founding teams: Homophily, strong ties, and isolation among U.S. entrepreneurs. American Sociological Review, 68(2), 195-222.
Shane, S., & Cable, D. (2002). Network ties, reputation, and the financing of new ventures. Management Science, 48(3), 364-381.
Unger, J. M., Rauch, A., Frese, M., & Rosenbusch, N. (2011). Human capital and entrepreneurial success: A meta-analytical review. Journal of Business Venturing, 26(3), 341-358.
Xiao, Y., Dowejko, M. K., Au, K., & Hsu, A. J. (2020). “Jack-of-all-trades” with passion: Keener to pursue startup in a team?. Journal of Small Business Management, 58(4), 806-833.

Appendix

Table A 1: Analysis of Gini coefficient and Jaccard index for both founders and teams of companies in different sectors and with different fundraising profile

				Outcome 0		Outcome 1		Outcome 2	
Label	Sector	#founders	#teams	G_N	J	G_N	J	G_N	J
1	aerospace	33	11	0.60	0.27	0.37	0.41	0.39	0.42
2	agriculture and farming	50	19	0.38	0.43	0.43	0.32	0.50	0.52
3	apps	414	158	0.41	0.35	0.42	0.41	0.41	0.34
4	as a service	557	209	0.43	0.33	0.40	0.32	0.43	0.36
5	augmented / virtual reality	92	34	0.45	0.32	0.43	0.34	0.34	0.23
6	beauty and wellness	159	60	0.43	0.33	0.42	0.28	0.45	0.22
7	clothing and apparel	136	45	0.39	0.27	0.37	0.33	0.42	0.23
8	cloud	222	78	0.36	0.42	0.40	0.37	0.37	0.36
9	commerce and shopping	576	192	0.40	0.34	0.40	0.33	0.41	0.38
10	communication protocols	56	17	0.48	0.36	0.41	0.42	0.32	0.58
11	community and lifestyle	196	77	0.47	0.27	0.55	0.28	0.45	0.32
12	consumer electronics	677	241	0.43	0.35	0.35	0.34	0.42	0.32
13	consumer goods	101	37	0.44	0.31	0.39	0.30	0.40	0.38
14	corporate applications	73	26	0.45	0.40	0.36	0.27	0.48	0.33
15	corporate functions	249	91	0.38	0.43	0.41	0.31	0.42	0.32
16	data and analytics	1029	370	0.44	0.36	0.46	0.37	0.43	0.32
17	design	149	53	0.39	0.39	0.25	0.33	0.43	0.22
18	development tools	198	81	0.45	0.38	0.45	0.44	0.49	0.33
19	education	234	79	0.39	0.35	0.37	0.30	0.41	0.37
20	electronic hardware	55	21	0.35	0.41	0.36	0.45	0.34	0.38
21	energy	96	36	0.39	0.35	0.44	0.38	0.42	0.29
22	events	81	31	0.48	0.29	0.50	0.35	0.38	0.45
23	financial services	556	196	0.40	0.33	0.42	0.37	0.49	0.39
24	food and beverage	163	56	0.37	0.32	0.28	0.40	0.42	0.42
25	gaming	83	35	0.47	0.32	0.36	0.23	0.41	0.31

26	government and military	57	19	0.37	0.31	0.45	0.20	0.42	0.29
27	hardware	101	33	0.43	0.36	0.40	0.37	0.40	0.33
28	health care	589	216	0.39	0.36	0.44	0.35	0.43	0.29
29	ict	705	251	0.41	0.38	0.46	0.32	0.40	0.37
30	internet of things	109	40	0.50	0.37	0.47	0.43	0.41	0.37
31	internet services	624	219	0.43	0.35	0.37	0.35	0.39	0.35
32	life science / biotech	171	60	0.39	0.35	0.45	0.48	0.36	0.35
33	manufacturing	163	59	0.43	0.27	0.36	0.34	0.39	0.42
34	media contents manag.	75	33	0.53	0.38	0.43	0.31	0.60	0.35
35	media digital media	86	34	0.41	0.32	0.39	0.33	0.27	0.35
36	media entertainment	100	36	0.45	0.35	0.41	0.29	0.37	0.42
37	media music and audio	65	26	0.45	0.36	0.41	0.39	0.40	0.17
38	media publishing	100	41	0.37	0.38	0.35	0.33	0.40	0.53
39	media social media	183	69	0.44	0.31	0.39	0.35	0.49	0.43
40	media video	128	47	0.45	0.37	0.30	0.53	0.34	0.39
41	messaging and telecom	154	52	0.44	0.29	0.37	0.38	0.41	0.42
42	navigation and mapping	55	19	0.44	0.28	0.42	0.48	0.38	0.22
43	platforms software	95	36	0.45	0.37	0.40	0.27	0.25	0.22
44	privacy and security	206	71	0.47	0.42	0.41	0.37	0.37	0.49
45	productivity	77	27	0.42	0.29	0.49	0.19	0.42	0.44
46	professional services	260	102	0.38	0.42	0.40	0.27	0.35	0.35
47	real estate	287	100	0.36	0.31	0.51	0.32	0.44	0.29
48	robotics	110	43	0.53	0.40	0.37	0.32	0.47	0.22
49	sales and marketing	663	249	0.43	0.37	0.39	0.36	0.42	0.34
50	software	1634	587	0.40	0.36	0.42	0.35	0.41	0.37
51	sports	178	64	0.39	0.31	0.36	0.40	0.38	0.24
52	sustainability	73	26	0.36	0.37	0.41	0.29	0.31	0.32
53	transportation	61	23	0.47	0.34	0.55	0.38	0.33	0.54
54	personal mobility	34	11	0.61	0.36	0.47	0.46	0.37	0.34
55	shipping and delivery	86	28	0.36	0.48	0.46	0.45	0.35	0.35
56	vehicles	89	30	0.42	0.40	0.34	0.29	0.32	0.28
57	travel and tourism	131	49	0.46	0.36	0.39	0.30	0.31	0.27
			Min	0.35	0.27	0.22	0.00	0.25	0.17
			Max	0.61	0.48	0.78	0.53	0.60	0.58
			Mean	0.43	0.35	0.41	0.34	0.40	0.35