



Research Article

Economic Interplay Forecasting Business Success

Nicola Amoroso ^{1,2}, **Loredana Bellantuono**,³ **Alfonso Monaco** ², **Francesco De Nicolò**,³
Ernesto Somma,^{4,5} and **Roberto Bellotti**^{2,3}

¹Università di Bari, Dipartimento di Farmacia-Scienze del Farmaco, Bari, Italy

²Istituto Nazionale di Fisica Nucleare, Sez. di Bari, Bari, Italy

³Università di Bari, Dipartimento Interateneo di Fisica, Bari, Italy

⁴Università di Bari, Dipartimento di Economia, Management e Diritto dell'Impresa, Bari, Italy

⁵Invitalia-Agenzia Nazionale per l'Attrazione Degli Investimenti e lo sviluppo d'impresa S.p.A, Roma, Italy

Correspondence should be addressed to Alfonso Monaco; alfonso.monaco@ba.infn.it

Received 11 August 2020; Revised 17 February 2021; Accepted 1 March 2021; Published 19 March 2021

Academic Editor: Giulio Cimini

Copyright © 2021 Nicola Amoroso et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A startup ecosystem is a dynamic environment in which several actors, such as investors, venture capitalists, angels, and facilitators, are the protagonists of a complex interplay. Most of these interactions involve the flow of capital whose size and direction help to map the intricate system of relationships. This quantity is also considered a good proxy of economic success. Given the complexity of such systems, it would be more desirable to supplement this information with other informative features, and a natural choice is to adopt mathematical measures. In this work, we will specifically consider network centrality measures, borrowed by network theory. In particular, using the largest publicly available dataset for startups, the Crunchbase dataset, we show how centrality measures highlight the importance of particular players, such as angels and accelerators, whose role could be underestimated by focusing on collected funds only. We also provide a quantitative criterion to establish which firms should be considered strategic and rank them. Finally, as funding is a widespread measure for success in economic settings, we investigate to which extent this measure is in agreement with network metrics; the model accurately forecasts which firms will receive the highest funding in future years.

1. Introduction

The economic interplay between firms and investors is of paramount importance in shaping the path and direction of the economic growth of a country [1]. There is an ongoing controversy about how growth and innovation are related, but not about the existence of such a connection [2–5]. Accordingly, there is an increasing interest in developing quantitative frameworks to identify and rate the strategic players of an economic ecosystem [6, 7], a task particularly difficult to accomplish when considering the dynamic and high-risk environment of startup companies [8, 9]. It is not a case that industrial innovation is one of the 17 Sustainable Development Goals promoted by the United Nations for the next decade; in fact, this aspect is crucial when considering growing economies and developing countries [10–12].

These considerations are gaining more and more ground so that recently some scholars have introduced the concept of *high-impact entrepreneurship* just thinking about startup firms [13, 14]. Their complex ecosystem, including moneylenders, investors, angels, banks, and financing agents, can help our understanding of the state of health of the economy. An old quote says: *Winning is not everything. It's the only thing*, but, what does winning mean in an economic setting? Is it right to have just one possible definition of *success* within such an intricate system of relationships? Are there many possible and complementary definitions?

Our analysis of complex economic ecosystems belongs to a research area, known as *science of success*, that is currently gaining considerable relevance [15]. This emerging sector of complex system analysis exploits the increasing availability of data to explore patterns that underlie success in diverse

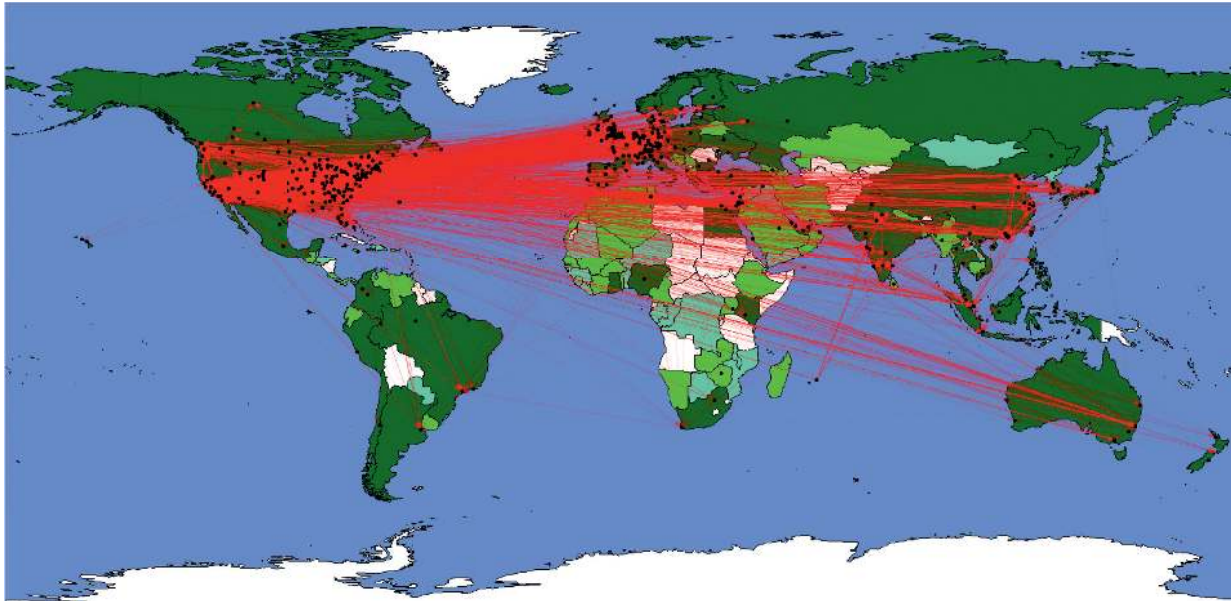


FIGURE 1: The Crunchbase network. A pictorial world map of the Crunchbase ecosystem. For each nation, the percentage of companies tracked within Crunchbase is represented. Nodes are in black and arrows in red.

areas, such as international country rankings [16], scientific publications [17], grant proposals [18], sports competitions [19], and patents [20]. The science of success investigates the impact of certain universal procedures such as partnership, mentoring, collaboration, or innovation, on the success of different initiatives, with the aim of identifying a number of common good practices, applicable in different contexts.

Inspired by the ideas of network success theory [21], we investigate the relation between the success of a startup, which a large body of literature defines as the capability of obtaining massive capital [22], and its capability to fully exploit its own business network. Actually, this attitude could shed light on complex economic dynamics [23, 24]. In particular, using a large public dataset, the proposed approach explicitly addresses the open questions raised by previous studies [25, 26], especially concerning the possibility of success being a direct consequence of a firm's networking.

As far as we know, few studies have investigated the economic systems of startup firms within quantitative frameworks [9]. Here, we present a quantitative assessment of interactions involving startups and their investors and a series of practical measurements borrowed by the mathematical graph theory to determine which agents, investors or beneficiaries, can properly be considered as "strategic" for the economic system under investigation.

One question addressed by our work is whether economic interplay can be accurately modeled with a complex network, thus providing an objective framework to define *strategic* actors within an economic system. Preliminarily, we demonstrate that the informative content provided by classical approaches, as those based on merely statistical methods, fails to capture the whole picture. Actually, the overall funding fails to fully identify actors playing key roles in the startup ecosystem. Funding does not represent how

many investors are involved in a funding round. As a matter of fact, the funds collected by a firm do not indicate its role in the setting, for example, it does not yield any information about which firms are connected to it, if it plays as an investor or it prefers the money conveyance.

To address this issue, we investigate here Crunchbase, a platform collecting a large amount of data on the startup ecosystem, with a special focus on investors, incubators, key-people, funds, funding rounds, and events. The Crunchbase dataset was created in 2007 by the TechCrunch Company, which managed it until 2015, when the Crunchbase platform became a private entity. According to OECD, these data have been used for over 90 scientific publications [27], whose subjects range from business administration and economics, with particular attention to venture capital and startup companies [28, 29], to psychological evaluations of entrepreneurship [30] and administrative science [31]. A particular mention must be given to studies concerning mathematical models, especially inspired by complex network approaches [32].

In fact, network theory is an extremely efficient tool to model complex systems, especially to highlight the importance of particular elements that in network jargon are called nodes. In this work, we investigate a network model whose nodes are the elements of the startup ecosystem: firms and investors; a directed edge is drawn between two firms if a funding relation holds, the origin being the investor firm. A basic idea to measure nodal importance is by means of its position within the network; the more a node is central, the more it is relevant. By extension, all measures trying to capture the importance of a node are called centrality measures, even when they have no direct geometric interpretation. In particular, three types of nodal importance measures can be distinguished, according to the way in which a node can influence the others [33, 34]: measures of immediate effects such as *degree*, measures of

mediative effects such as *betweenness*, and measures of total effects such as *eigenvector centrality*. In this work, we consider strategic those firms whose behavior in terms of funds, degree, or betweenness significantly differs from other firms. The measures expressed by network centralities represent information complementary to that provided by collected funds and help to highlight different points of view on the startup ecosystem. For example, degree centrality measures the overall number of connections of a node, and the underlying assumption is that the larger the number of connections, the greater the importance of the node. The degree distribution unveils important properties of a network, like the scale-free structure [35, 36]. Another example is the betweenness centrality [37]; this measure evaluates the node importance by taking into account the number of paths within a network exploiting that specific node; in this sense, it depicts the nodal importance with a more dynamic flavor. For the last class of measures, we can mention eigenvector centrality [38]. However, we will not further investigate this kind of centrality, since it is typically employed to characterize nodes in undirected networks, while, as we shall discuss in the following, our study is based on a complex network model with directed edges.

Accordingly, this work explores the use of degree and betweenness to highlight the different roles played by economic actors and rank their importance. The proposed analysis reveals information about the firms which could not be retrieved otherwise by merely inspecting the collected funds. Nonetheless, as collected funds are widely adopted as a measure of *success*, in the second part of this work, we investigate how network centralities can be seen as a proxy of future *success* defined as being an outlier in the distribution of collected funds. In particular, our model relates the network centralities of a firm to its possibility of being a funding outlier in a future time.

2. Materials and Methods

2.1. Crunchbase Data. In this study, we focus on the Crunchbase dataset, a huge set of data collected on the crunchbase.com site. Specifically, our results are based on the 13 October 2017 update. This site hosts contributions from all over the world and is, to date, widely considered as one of the most comprehensive publicly available datasets about investment and funding on a global scale, as it contains more than 50 million records. More precisely, Crunchbase includes detailed information on more than 550000 companies from 160 countries (Figure 1), distributed among 38 different economic categories. Nonetheless, it is worth emphasizing that not all the companies and investors contained in the dataset are involved in funding rounds, but 121950 of them actually are. These latter elements are of interest for our subsequent analysis.

Some of these firms are investors, classified into 10 possible types. Crunchbase data are organized in 17 distinct datasets, as listed in Table S1 in the Appendix and focusing on several specific subjects, such as acquisitions, economic categories, collected funds, personnel, investment partners, and geographic site. Besides, Crunchbase includes information about funding events, such as how many funders are

involved, how much money (in USD) was collected in a funding event, and its date. In particular, funds are reported back to 1960.

By use of the data available on Crunchbase, we are able to accurately track the flow and direction of investments and identify those companies (VCs, startups, and business angels) that outperformed in attracting and/or investing capital. Besides, we took into account geographical information about the firms to geolocalize investment patterns, the economic category describing the business activity of each company, and the investor role (e.g., angel and accelerator) played by the various agents within the economic system.

Crunchbase companies are almost ubiquitous; nevertheless, the USA is by far the leading country (53.6%), as expected being the USA an extremely favorable country for this kind of business; it is worth noting that the second country is the UK with only 7.6%. Among different economic categories present in Crunchbase, Internet services and e-Payments play a leading role accounting for 19.3% and 14.4%, respectively; software (6.1%), science (5.8%), and ICT (5.6%) firms have also a consistent representation. Finally, concerning the investor types, the most frequent ones are angels (60%) and venture capitalists (28%), while other categories have occurrences not exceeding 5% (Table S2).

2.2. Modeling the Economic Interplay. Based on information about the flow of investments, we modeled this economic interplay with a directed complex network: nodes represent all the elements reported in Crunchbase, both startups and funders, while the directed links correspond to the investments, the origin being the investor company (funder) and the end being the one receiving funds. The reason for such a model is twofold: on the one hand, we get a representation adherent to traditional economic approaches monitoring the money flux; on the other hand, this model of economic interplay is straightforward and easy to interpret. A not secondary aspect to mention is that, thanks to this model, we can provide a quantitative evaluation of nodal importance. Thus, we can establish to which extent a firm plays a strategic role within the economic system and measure the success probability of its business.

We denote as N the set of Crunchbase economic players and L the set of past economic transactions, so that, for each pair of nodes $n_i, n_j \in N$, a transaction $(i, j) \in L$ is a flux of money from n_i to n_j . Accordingly, the directed graph G , denoted as the couple (N, L) , has order $|N| = 121950$ and size $|L| = 289396$. Of course, this graph is not symmetric as the existence of a connection (i, j) does not imply the existence of its counterpart (j, i) . It is worth noting that the network model is built using all transactions occurred between 1960 and October 2017.

Crunchbase does not keep track of the amount of each transaction, so that a weighted description of the graph is not accessible. Nevertheless, we do know the overall amount of collected funds for each company. Considering the amount of collected funds f as a proxy variable of the business success of each company and given the country c , the economic category e , and the investor type t as auxiliary

attributes, each node can be parametrized as $n_i = n_i(f; c, e, t)$. The primary goal of this work deals with investigating the existence of significant relationships among these four variables, four distinct assets for a successful firm.

Even if f is a fundamental measure of nodal importance, in this work, we demonstrate that it does not yield an exhaustive picture of the economic system under analysis; on the contrary, the network properties assessing the flux of capitals can result in a significant improvement of its description. Complex network theory provides us with several mathematical tools to evaluate this aspect. We consider three centrality metrics for each node n_i , namely, the *indegree*

$$k_i^{(\text{in})} = \sum_{j \neq i}^N l_{ji}, \quad (1)$$

where l_{ji} is 1 if there is a link incident on node n_i from node n_j and 0 otherwise, the *outdegree*

$$k_i^{(\text{out})} = \sum_{j \neq i}^N l_{ij}, \quad (2)$$

and the *betweenness*

$$b_i = \sum_{s,t} \frac{g_{st}^{(i)}}{g_{st}}, \quad (3)$$

where $g_{st}^{(i)}$ denotes the number of geodesic paths starting from node n_s , passing through the node n_i and reaching the node n_t , and g_{st} denotes the cardinality of the whole set of geodesics starting from n_s and ending to n_t . We chose these three measurements as the most suitable ones to characterize some specific properties of companies in Crunchbase:

- (1) The investor attractiveness
- (2) The financing power
- (3) The capital conveyance

Capturing this information provides a deeper knowledge on the economic system of startup firms, as it takes into account how funds are collected, outsourced, and conveyed. It is worth noting that, in general, there is no reason for these three distinct actions to be performed by the same agent; on the contrary, it is reasonable to assume that, according to each aspect, specific strategic actors can be identified. Besides, from this picture, it is also manifest that considering only the amount of funds collected by a firm provides too limited a description of the system.

2.3. Defining and Measuring Success. A straightforward definition of success for a startup business, at least until it becomes profitable, is the amount of capital it is able to collect. This definition seems reasonable in terms of both meaningfulness and interpretability; another key aspect is that capitals are quantitatively measurable, and thus, they provide an objective strategy to evaluate success.

The amount of funds collected by a startup is a reliable measure of its success, but provides a limited picture of what happens in the startup ecosystem. For example, the amount

of collected funds does not contain information about the number of funders and obviously does not quantify the capability or willingness to fund other firms, as well as the attitude to convey capitals within the system.

To answer these questions, a richer set of information about the system should be taken into account, instead. For example, within an economic system, there are companies whose main role is not that of collecting capitals, but investing them. Accordingly, their importance would be hidden if considering only the amount of collected funds; nevertheless, their presence is an invaluable asset for business. Another crucial aspect concerns the way capital moves throughout the economic system. In network theory, it is well known that some nodes can deeply influence other nodes even when they are not directly connected, but thanks to an indirect influence. Moreover, a comprehensive description of the startup ecosystem should distinguish cases in which subjects collect similar amounts of funding but employ them in very different ways, e.g., for their own expenses or to finance other firms.

We investigated the distribution of collected funds within Crunchbase and compared it with the distributions of indegree, outdegree, and betweenness. By applying the nonparametric Kolmogorov–Smirnov test, we found a statistically significant difference ($p < 10^{-3}$) between all centrality distributions and the funding one (Figure S1). This analysis confirmed that the informative content provided by network centralities does not significantly overlap with that provided by funds. Then, for each distribution, we determined the outlier observations. The distributions considered in this work are positive definite; in fact, both funds and network centralities admit only positive values. We hypothesized here that *strategic* companies are simply the right outliers, as they were able to collect funds, investors, investments, and capital transfers significantly better than others. The outliers are defined as those elements with high values of centrality measures (funding, indegree, outdegree, and betweenness), thus obtaining four kinds of outliers. Since we need a quantitative definition of *high values*, a standard procedure to define the outliers is used: the boxplot method. For each centrality measure, all the elements whose values exceed the threshold value given by the 75th percentile of their distribution added to $1.5 \times$ the interquartile range (IQR) are defined to be an outlier. In this sense, they are *successful* companies, and further methodological details are provided in Appendix S1.

3. Results and Discussion

3.1. Successful Companies. We found 7176 outliers for the distribution of funding, 14716 for indegree, 12846 for outdegree, and 1523 for betweenness. Besides the bare numeric differences, which demonstrate how the number of strategic elements strongly depends on the definition of importance adopted, further insights were obtained by computing the Kendall correlation τ between each centrality distribution and the funding one. Results reveal that the indegree centrality has the highest correlation with the amount of collected funds ($\tau = 0.4$), a rather intuitive

TABLE 1: Best performers for nationality, category, and type.

	Fund	Indegree	Outdegree	Betweenness
<i>Nationality</i>				
(1)	USA (70%)	USA (72%)	USA (57%)	USA (55%)
(2)	CHN (7%)	UK (5%)	UK (7%)	CHN (9%)
(3)	UK (4%)	CAN (3%)	DEU (3%)	IND (5%)
<i>Category</i>				
(1)	Is (21%)	Is (25%)	eP (72%)	eP (24%)
(2)	Sc (15%)	Sc (9%)	Is (6%)	Is (21%)
(3)	eP (8%)	eP (8%)	Sc (2%)	Sc (7%)
<i>Type</i>				
(1)	VC (63%)	VC (54%)	VC (50%)	VC (51%)
(2)	PE (15%)	Acc (35%)	Ang (31%)	Acc (25%)
(3)	Inv (4%)	HF (4%)	PE (9%)	PE (10%)

Comparison of top three rankings according to funding (Fund), indegree (In), outdegree (Out), and betweenness (Bet). Nations are abbreviated according to International Naming Convention. Categories: Internet services (Is), e-Payments (eP), and Science (Sc). Investor types: venture capital (VC), private equity (PE), accelerators (Acc), angels (Ang), investment bank (Inv), and hedge funds (HF).

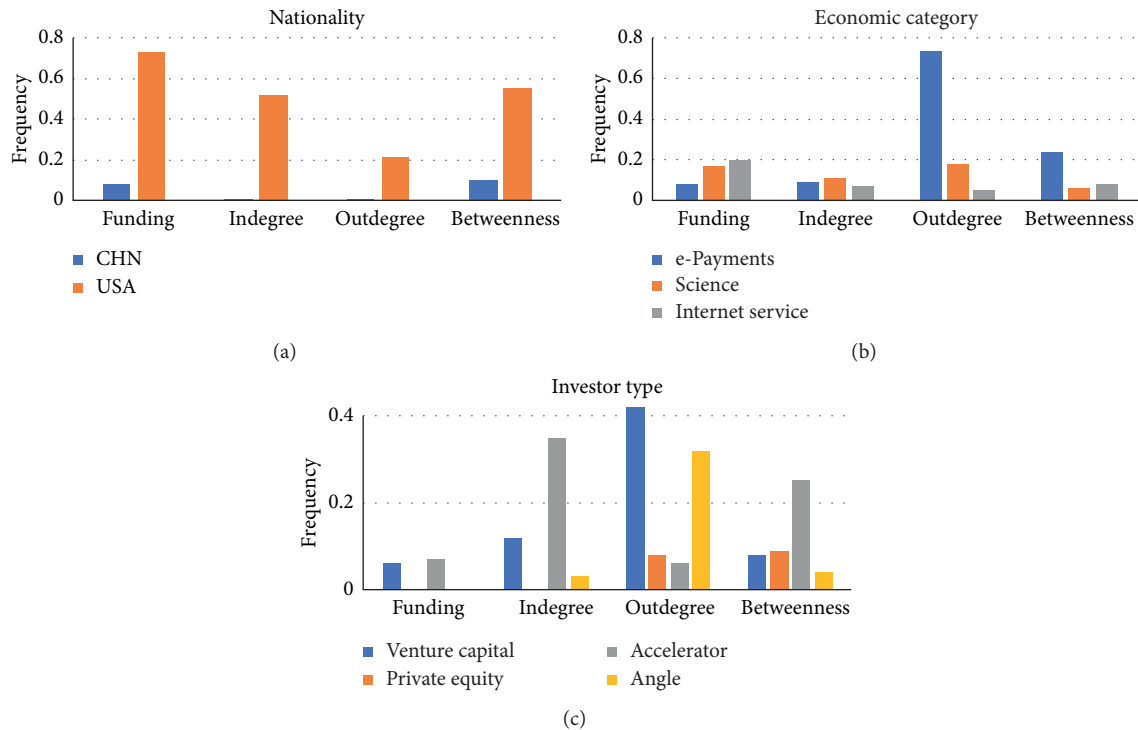


FIGURE 2: Top performers. (a) Frequencies of nationalities, (b) economic categories, and (c) investor types in the outliers of funding and centrality metrics distributions.

outcome, while the outdegree and betweenness are less correlated ($\tau = 0.1$ for both of them). All three correlations have a 1% statistical significance, which demonstrates the existence of monotone relationships between funds and centrality measures. The top 50 firms for each ranking are alleged in Table S3; a synthetic overview is presented in Table 1.

These findings, somehow expected for what concerns outdegree and betweenness, could appear, at first glance, surprising for indegree. In fact, it is reasonable that the ability of a firm to collect funds should be proportional to the number of investors it can relate with. On the contrary, this

result suggests that large investments tend to arrive in solitude and that when a firm is able to collect funds from different sources, it is probable these will be small funds. Nevertheless, this is not the only conclusion we can draw from this model; a further characterization, which the previous rankings cannot outline, can be provided instead in terms of economic categories and investor types.

Of course, success is a multifaceted concept and can be defined in many alternative ways, e.g., by using profitability measures such as the income flow or by considering startup acquisition and initial public offering (IPO). However, these aspects fall outside our scope and represent complementary

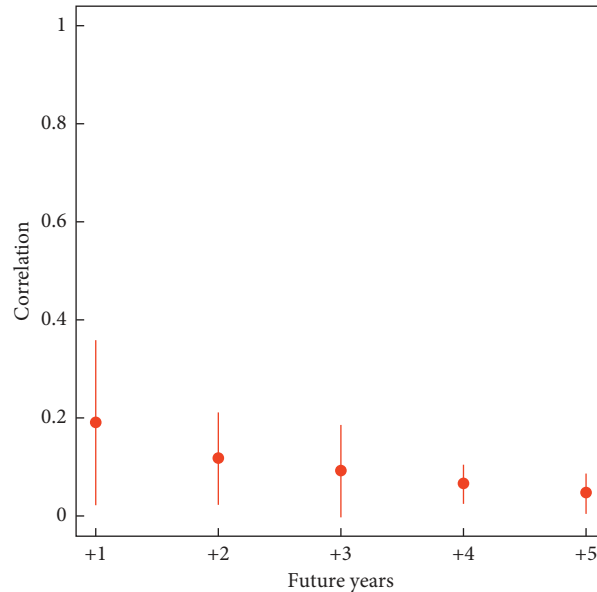


FIGURE 3: Mean correlation. Mean correlation variation, along with its standard deviation, with future years.

viewpoint, with their own peculiarities and interpretation difficulties, in characterizing the startup system. In conclusion, the choice to consider successful firms according to collected funds yields is twofold: (i) it is intuitive and (ii) widespread in economic literature [21, 22].

3.2. Investors and Economic Categories. The sole inspection of funding outliers unveils important information about success. The results on top nations, economic categories, and investor types, reported in Table 1, confirm what has been found in other studies [39, 40], with different data. What can we say for network centralities? Do they either confirm these findings or provide novel insight? To answer these questions, we compared the funding outliers with the indegree, outdegree, and betweenness ones and found significant differences ($p < 0.05$ Bonferroni-corrected) for nationalities, economic categories, and investor types. In particular, our analyses highlighted the role played by the USA and Chinese firms for what concerns nationality; e-Payments, Science, and Internet services for economic category; and finally, venture capital, private equity, accelerator, and angel for investor type (Figure 2).

Further details about this analysis are presented in Figures S2(a)–S2(c), S3(a)–S3(c), and S4(a)–S4(c). USA firms are able to collect more funds than expected just looking at network centralities, the larger difference being between funding and outdegree; this is not surprising as the USA hosts the majority of Crunchbase firms and provides extremely advantageous economic conditions, especially for startups. It is instead surprising that the prevalence of USA firms among outdegree outliers is much smaller (around 20%) than for the other distributions. Of course, the fact that USA firms are the most frequent among the Crunchbase elements importantly affects these results; nevertheless, the fact that a country is present with a given frequency does not

ensure that its attributes (funding, indegree, outdegree, and betweenness) should be outliers with the same frequency. For example, for what concerns nationality, Figure 2 shows that this happens only for USA and China. In these nations, we observe a significant difference between the frequency of funding outliers and network outliers. In particular, USA firms are able to collect more funds than expected just looking at network centralities, the larger difference being between funding and outdegree.

The startup ecosystem encompasses almost entirely the whole range of economic sectors; through the examination of how funds are distributed among successful firms, we established that Science applications and Internet services are generally the economic categories able to collect the largest amounts of funds. In fact, these two categories account together for about 38% of funding outliers. On the contrary, network centralities, especially outdegree and betweenness, outline the role played by e-Payments. Actually, e-Payment firms represent 72% of outdegree outliers and 23% of betweenness outliers, a result which makes sense as this specific economic sector is particularly devoted to capital investments and conveyance.

Finally, for what concerns investors, we found 4 significant outcomes: (i) Venture capital firms have an outstanding presence among outdegree outliers, according to their compelling vocation for investments. (ii) Private equities show a significant presence among outdegree and betweenness outliers; on the contrary, they are completely absent from indegree and funding outliers, suggesting their strategic role in investments and capital conveyance. (iii) We observed a significantly larger presence of accelerators among indegree and betweenness outliers, suggesting an interesting interpretation: strategic accelerators are oriented to collect funds of small/medium entity from a large number of investors and convey them to other firms. Thus, they are strategic players, acting as connection hubs within the

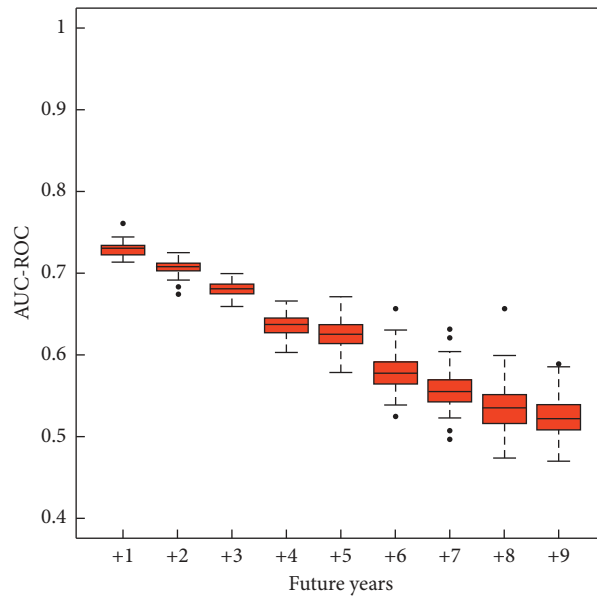


FIGURE 4: Forecasting success. Area under the curve (AUC) of predictions up to ten years obtained with the logistic regression model.

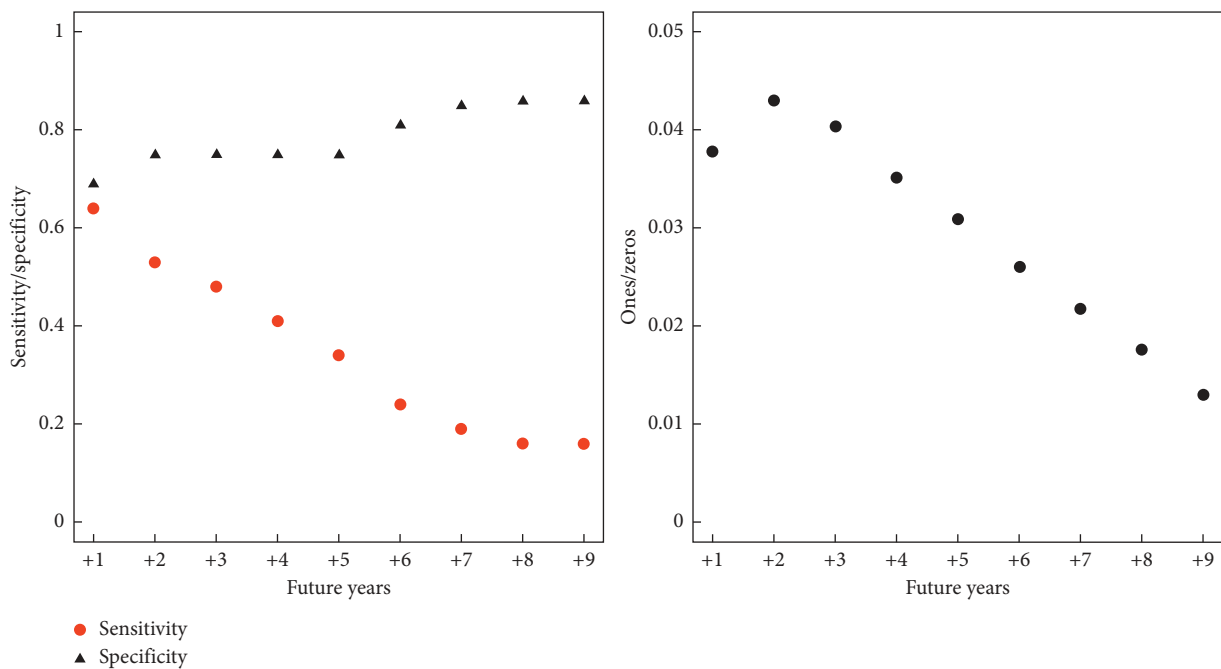


FIGURE 5: Sensitivity and specificity analysis. (a) Sensitivity and specificity of results of the model and (b) its class ratio.

startup ecosystem. Their crucial role in the network structure would have been neglected in an analysis based only on funding outliers, a set in which accelerators represent only 7% of firms, while their frequencies among indegree and betweenness outliers are 38% and 24%, respectively. (iv) The outdegree outliers show a significantly larger presence of angels (36%) compared with other distributions, a result outlining the fundamental role played by these investors in granting funds to a large number of firms. Even in this case, this role would not be noticed by just looking at the funding distribution, where angels do not appear at all.

3.3. Forecasting Success. So far, we have essentially outlined two different points: firstly, the informative content provided by funding is significantly different from that provided by network centrality measures; secondly, funding and centrality characterize different strategic aspects of an economic ecosystem. Now, we address the two last questions. Identifying successful firms with the outliers of funding distribution, are network centralities proxies of this notion of economic success? If yes, to which extent?

Provided that in Crunchbase each firm is a node $n = n(f; c, e, t)$, we investigated to which extent we could

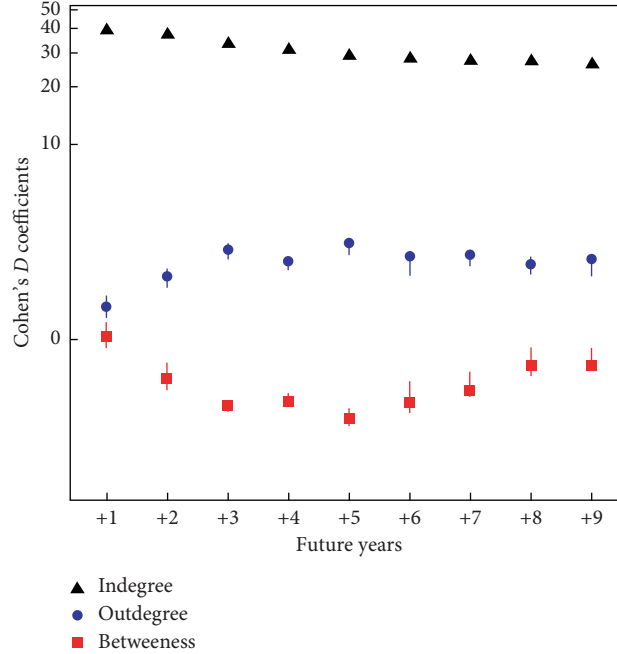


FIGURE 6: Effect size. For each year, the importance of node centralities is expressed in terms of Cohen's D coefficients in a symmetric log scale.

formulate an alternative description by modeling funding where i , o , and b are the proposed centrality measures: indegree, outdegree, and betweenness, respectively. It is worth noting that, based on the peculiar nature of the startup funding (which is usually a one-time-event), the amount of collected funds in one funding round is weakly correlated to those raised in successive funding rounds (see Figure 3).

The figure shows how correlation is rather weak even at low values of future years (0.2 at 1-future year), approaching zero as the time interval between the two observations increases.

Multiple supervised strategies could be applied; however, for the sake of interpretability and given the exiguous number of independent predictor variables, we chose a logistic regression approach [41]. Formally, our outcome variable f is 1 for a successful firm and 0 otherwise; we express f as a function of (i, o, b) :

$$f = \frac{e^{\beta_0 + \beta_1 i + \beta_2 o + \beta_3 b}}{1 + e^{\beta_0 + \beta_1 i + \beta_2 o + \beta_3 b}}. \quad (4)$$

It is worth noting that Crunchbase observations date back to 1960; however, until 1999, only 2739 records were acquired; they were 10221 just considering the year 2000. Accordingly, to forecast business success, we considered only data collected from 2000 to 2017, thus resulting in 78298 firms. Besides, when considering forecasting, we restricted our data to firms surviving up to 9 years. For each year y and for each node n_j , we evaluated the indegree $i_j^{(y)}$, the outdegree $o_j^{(y)}$, and the betweenness $b_j^{(y)}$, which are the independent variables of the model. The dependent variable $f_j^{(y)}$ indicates whether node n_j in year y is an outlier for collected funds or not.

For every year $T \in \{2000, \dots, 2017\}$ in the dataset, we built the related network and computed the nodal centralities; then, for each node, we determined if in a future year it corresponded or not to a funding outlier; successful firms were labeled with $\mathbf{1}$ and $\mathbf{0}$ otherwise. Then, we trained a model at a time T and used the future years for test. The analysis was carried out within a 5-fold cross-validation framework, and the procedure was repeated 100 times. Finally, we used network centralities to predict whether after 1, 2, \dots years a firm will be a funding outlier and evaluated the performance of the model in terms of the area under the receiver operating characteristic curve (AUC); results are shown in Figure 4.

These results show a significant association between network centralities and the amount of collected funds up to four/five years in the future, with median AUCs ranging from 0.73 (+1 year) to 0.61 (+5 years). As expected, the forecasting accuracy decreases as we move forward in time; the prediction to 9 years is barely distinguishable from random. Our findings emphasize the robustness of predictions, as training and validation performances do not show significant differences.

Besides, we examined sensitivity and specificity and their variation according to the ratio between successful and unsuccessful firms for each year (see Figure 5).

Two considerations arise: (i) our model's ability of retrieving nonfunding outliers (specificity) slightly grows over time and (ii) the performance drop observed in terms of AUC values is caused by the worsening of sensitivity, i.e., the capability to detect successful firms. This effect is dominated by the substantial drop of these firms over time; in fact, the successful firms which initially represented 4/5% of the data after 9 years were only 1%.

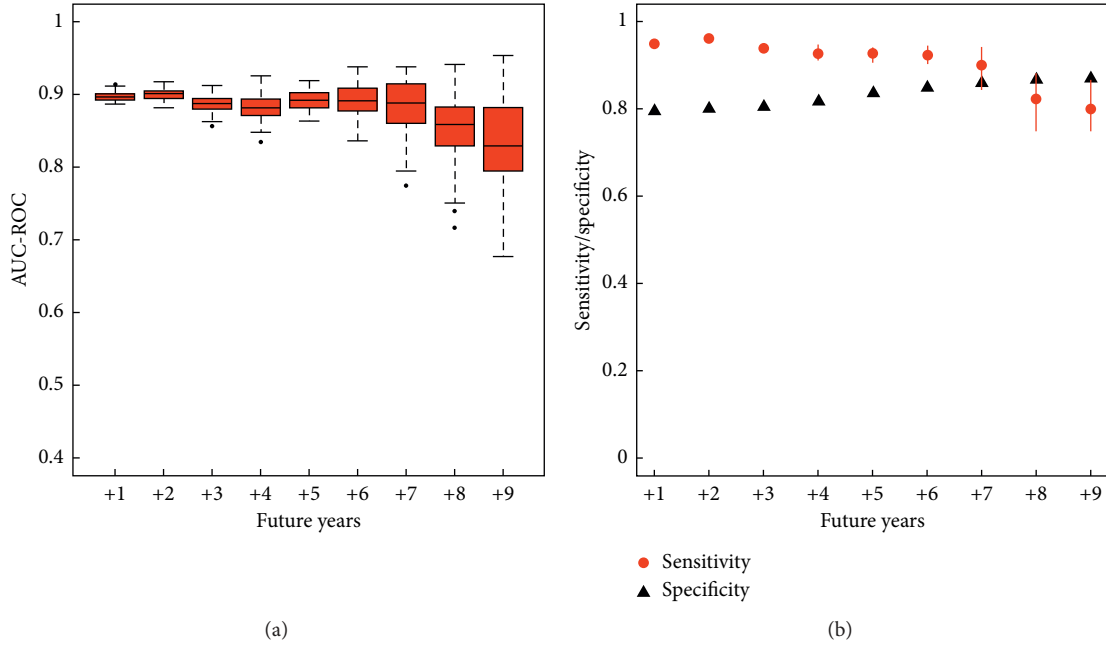


FIGURE 7: AUC-ROC, sensitivity, and specificity for the logistic regression model with network centralities only for the “e-Commerce” coarse economic category.

TABLE 2: The effect of present funding and firm-level variables on the model accuracy.

Year	Original model AUC/sens/spec (%)	Present funding AUC/sens/spec (%)	Firm-level variables AUC/sens/spec (%)
+1	73/64/69	83*/56*/86*	81*/79*/68
+2	71/53/75	79*/49*/86*	81*/80*/68*
+3	67/48/75	73*/40*/85*	80*/79*/68*
+4	63/41/75	67*/32*/85*	79*/73*/69*
+5	62/34/78	63/23*/87*	77*/69*/71*
+6	58/24/81	59/16*/88*	73*/55*/74*
+7	56/19/85	57/15*/88*	71*/48*/77*
+8	54/16/86	54/14*/89*	69*/42*/79*
+9	53/16/86	52/12*/89*	68*/36*/81*

These values refer to the variance of the median values of the yearly distributions of AUC-ROC, sensitivity, and specificity. Significant differences according to the Mann–Whitney test at 1% significance level with a Bonferroni correction are marked (“*”).

To evaluate the importance of the different predictors, we used Cohen’s D [42]. Cohen’s D is an effect size measure; it compares the difference of two sets of observations or measures with their intrinsic variability:

$$D = \frac{E[X_1] - E[X_2]}{\sigma}, \quad (5)$$

where $E[X_1]$ and $E[X_2]$ denote the expectation values for the sets of observations X_1 and X_2 , respectively, and σ is the standard deviation divided by the square root of the number of training observations. Using Cohen’s D to evaluate the feature importance in the logistic regression, we found that the indegree is the most relevant feature to predict success in collecting funds (see Figure 6).

This result is particularly evident at very short time ranges (+1 year); interestingly, at time scales between +1 year and +3 years, the effects of both outdegree and

betweenness increase. For larger times, the indegree maintains its paramount importance while the other centralities remain comparable, but with different signs. These results would suggest that, in the *long period*, the successful firms are not only those able to collect capitals from many investors, but also those playing an active role in financing activities; negative coefficients in a logistic regression model translate into odds ratios that are less than one. This, in turn, means that the predicted probability is decreasing as the covariate increases. Thus, conveying capitals would seem instead anticorrelated to success. Finally, collecting capitals is an important asset for economic interplay, but its impact yields a greater effect at a short time range.

Interestingly, the more an element is able to facilitate the money flux in the startup ecosystem, the more its probability of having success in future years decreases. This result clearly indicates that, even if money conveyance can

be considered an asset [25, 26], it should be considered with caution when collecting funds. However, it should be taken into account that startup funding is the only funding mechanism considered here. Many betweenness outliers are stable and powerful firms (e.g., Alibaba, Google, Yahoo, Amazon, and Uber) which obviously do not focus their activities on collecting funds in the examined startup ecosystem, but do have an important role as publicly acknowledged mentors, thus justifying their prominent role in conveying money.

A primary goal of this work was to provide an interpretable model; accordingly, we adopted here a logistic regression approach. In particular, the logistic regression has manifest advantages: it returns both a measure of importance for each predictor, given by the magnitude of coefficients, and the direction of association, namely, the sign of coefficients. Nonetheless, other learning and modeling strategies (e.g., random forests, deep learning, and multiplex networks) could be adopted and could represent an interesting theme for future works.

In order to complete the characterization of the predictive power of the model, we apply the classification algorithm to detect the funding outliers in each of the 8 coarse economic categories, defined in Appendix S2 (see Table S4). Results are shown in Figures S5(a)–S5(d) and S6(a)–S6(d) in Supplementary Material. In order to compare the performances of the original model with those for each coarse economic category, we performed Kolmogorov–Smirnov test and the coarse category “e-Commerce” witnesses an improvement of all its performance measures (AUC-ROC, sensitivity, and specificity), while all the other coarse categories do not. Graphs for this coarse economic category are shown in Figure 7.

These findings suggest that the model provides reliable predictions concerning the success in funding round events of startups belonging to this coarse category (encompassing the original e-Commerce and e-Payments categories). The result is particularly interesting in view of the practical application of our model, since it proves to be able to predict funding success in such a growing and influential sector.

3.4. Insight: the Role of Funding and Firm Variables. Previous findings show how network centralities and funding correlate. We investigate now whether the funds collected at present time are a good proxy for future funds (for example, a trivial Matthew effect can be in action) and to what extent the introduction of other firm-level variables (e.g., economic category, investor type, and employee number) in the model can improve the model’s performance.

First of all, we assessed the effect of present funds on the forecast of future funds. We previously showed that the amount of funds collected within a specific year is weakly correlated to the amounts collected during the successive funding rounds. In addition, we considered a model including present funding as a predictor. For further assessment, we also included within the model

some firm-level variables. In particular, we considered (i) the investor type to distinguish the role played by each firm (e.g., angel and accelerator); (ii) the economic category to which a firm belongs; (iii) nationality of a firm to consider the geographical differences; and (iv) employee count to take into account a firm’s size. Each variable required specific considerations and preprocessing to be included in the model, and such details are presented in Appendix S2. The results of these analyses are summarized in Table 2.

Significant differences with respect of the original model are assessed with a Mann–Whitney test (1% significance), with a Bonferroni correction for multiple tests. The introduction of funding as a predictor involves a small but significant improvement, especially from +1 to +5 years. Interestingly, the inclusion of present funding lowers the model sensitivity, thus confirming the poor relationship between present and future funding. The model accuracy, with respect to all the three proposed metrics, significantly improves when enlarging the bucket of available predictors; the drawback is that such information can be often difficult to collect.

4. Conclusion

In this study, we developed a quantitative and easy-to-interpret model to account for the strategic importance of firms within an economic infrastructure. Last, but not least, we also demonstrated that, although funding and network centralities explain different effects, with a logistic regression model, it is possible to forecast the success of a firm up to five years in advance using only network metrics. Specifically, we identified the indegree as the most important centrality metrics to predict whether a firm is an outlier of the distribution of collected funds. Moreover, it is worth emphasizing that the logistic model using network centralities alone as predictors has excellent performances in predicting future success for elements belonging to the e-Commerce and e-Payments economic categories.

Finally, our study paves the way for future investigations, for example, about the existence of a relationship between the investor types and the economic categories or between the nationality of firms and their investors. The determinants of success for firms of different nationality, type, or category are likely to be different. A preliminary analysis is reported in this work by considering a few firm-level variables; however, it would be of paramount importance to further investigate this aspect and understand how different mix of factors can be successful.

Data Availability

The Crunchbase dataset is publicly available at <https://www.crunchbase.com/>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Supplementary Materials

In the Supplementary Material attached to this article, we first insert a graph reporting the funding and network centralities' distributions (indegree, outdegree, and betweenness) for all the elements in the Crunchbase complex network we build in the "Modeling the economic interplay" section. Note that this graph is in semilogarithmic scale. Moreover, we add a series of couple of graphs, from Figure S2(a) to Figure S4(c), comparing the attributes' frequencies for funding outliers to those for the outliers of each network centrality. In the end, from Figure S5(a) to Figure S6(d), we report a sequence of graphs representing performances of our logistic regression model (with network centralities only as predictors) applied to each of the coarse economic category, as explained in the "Forecasting success" section of this paper. (*Supplementary Materials*)

References

- [1] T. J. Kane, *The Importance of Startups in Job Creation and Job Destruction*, 2010.
- [2] T. Suri, M. A. Boozer, G. Ranis, and F. Stewart, "Paths to success: the relationship between human development and economic growth," *World Development*, vol. 39, no. 4, pp. 506–522, 2011.
- [3] R. Florida, "The creative class and economic development," *Economic Development Quarterly*, vol. 28, no. 3, pp. 196–205, 2014.
- [4] D. Acemoglu, U. Akcigit, and W. R. Kerr, "Innovation network," *Proceedings of the National Academy of Sciences*, vol. 113, no. 41, pp. 11483–11488, 2016.
- [5] N. Lee, "Psychology and the geography of innovation," *Economic Geography*, vol. 93, no. 2, pp. 106–130, 2017.
- [6] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero, "A new metrics for countries' fitness and products' complexity," *Scientific Reports*, vol. 2, p. 723, 2012.
- [7] P. Mealy, J. D. Farmer, and A. Teytelboym, "Interpreting economic complexity," *Science Advances*, vol. 5, no. 1, 2019.
- [8] Y. Yin, Y. Wang, J. A. Evans, and D. Wang, "Quantifying the dynamics of failure across science, startups and security," *Nature*, vol. 575, no. 7781, pp. 190–194, 2019.
- [9] M. Bonaventura, V. Ciotti, P. Panzarasa, S. Liverani, L. Lacasa, and V. Latora, "Predicting success in the worldwide start-up network," *Scientific Reports*, vol. 10, no. 1, pp. 1–6, 2020.
- [10] A. Mesnard and M. Ravallion, "The wealth effect on new business startups in a developing economy," *Economica*, vol. 73, no. 291, pp. 367–392, 2006.
- [11] R. Dellink, J. Chateau, E. Lanzi, and B. Magné, "Long-term economic growth projections in the shared socioeconomic pathways," *Global Environmental Change*, vol. 42, pp. 200–214, 2017.
- [12] J. C. Cuaresma, W. Fengler, H. Kharas, K. Bekhtiar, M. Brottrager, and M. Hofer, "Will the Sustainable Development Goals be fulfilled? Assessing present and future global poverty," *Palgrave Communications*, vol. 4, no. 1, pp. 1–8, 2018.
- [13] Z. Acs and D. Audretsch, *Handbook of Entrepreneurship Research: An Interdisciplinary Survey and Introduction*, Vol. 1, Springer Science & Business Media, New York, NY, USA, 2006.
- [14] S. J. Davis, R. J. Faberman, J. Haltiwanger, R. Jarmin, and J. Miranda, "Business volatility, job destruction, and unemployment," *American Economic Journal: Macroeconomics*, vol. 2, no. 2, pp. 259–287, 2010.
- [15] A. L. Barabasi, *The Formula: The Universal Laws of Success*, Little, Brown and Company, Hachette Book Group, New York, NY, USA, 2018.
- [16] L. Bellantuono, A. Monaco, S. Tangaro, N. Amoroso, V. Aquaro, and R. Bellotti, "An equity-oriented rethink of global rankings with complex networks mapping development," *Scientific Reports*, vol. 10, p. 18046, 2020.
- [17] A. Zeng, Z. Shen, J. Zhou et al., "The science of science: from the perspective of complex systems," *Physics Reports*, vol. 714–715, pp. 1–73, 2017.
- [18] S. A. Gallo, J. H. Sullivan, and S. R. Glisson, "The influence of peer reviewer expertise on the evaluation of research funding applications," *PLoS One*, vol. 11, no. 10, Article ID e0165147, 2016.
- [19] B. Gonçalves, D. Coutinho, S. Santos, C. Lago-Penas, S. Jiménez, and J. Sampaio, "Exploring team passing networks and player movement dynamics in youth association football," *PLoS One*, vol. 12, no. 1, Article ID e0171156, 2017.
- [20] P. Érdi, K. Makovi, Z. Somogyvári et al., "Prediction of emerging technologies based on analysis of the US patent citation network," *Scientometrics*, vol. 95, no. 1, pp. 225–242, 2013.
- [21] A. Dessyana, B. P. Dwi Riyanti, and B. P. Dwi Riyanti, "The influence of innovation and entrepreneurial self-efficacy to digital startup success," *International Research Journal of Business Studies*, vol. 10, no. 1, pp. 57–68, 2017.
- [22] P. Witt, "Entrepreneurs' networks and the success of startups," *Entrepreneurship and Regional Development: An International Journal*, vol. 16, no. 5, pp. 391–412, 1993.
- [23] E. L. Hansen, "Resource acquisitions as a start-up process: initial stocks of social capital and organizational foundations," *Frontiers of Entrepreneurship Research*, pp. 111–120, 2000.
- [24] G. N. Chandler and S. H. Hanks, "Measuring the performance of emerging businesses: a validation study," *Journal of Business Venturing*, vol. 8, no. 5, pp. 391–408, 1993.
- [25] G. Trp, *Financing High-Tech Startups*, Springer, New York, NY, USA, 2018.
- [26] A. Salamzadhe, "Start-up boom in an emerging market: a niche market approach," *Competitiveness in Emerging Markets*, pp. 233–234, 2018.
- [27] J. M. Dalle, M. Den Besten, and C. Menon, *Using Crunchbase for Economic and Managerial Research*, 2017.
- [28] O. Alexy, J. H. Block, P. G. Sandner, and A. L. Ter Wal, *The Social Capital of Venture Capitalists and its Impact on the Funding of Start-Up Firms*, 2010.
- [29] O. T. Alexy, J. H. Block, P. Sandner, and A. L. J. Ter Wal, "Social capital of venture capitalists and start-up funding," *Small Business Economics*, vol. 39, no. 4, pp. 835–851, 2012.
- [30] A. Tata, D. L. Martinez, D. Garcia, A. Oesch, and S. Brusoni, "The psycholinguistics of entrepreneurship," *Journal of Business Venturing Insights*, vol. 7, pp. 38–44, 2017.
- [31] A. L. J. Ter Wal, O. Alexy, J. Block, and P. G. Sandner, "The best of both worlds," *Administrative Science Quarterly*, vol. 61, no. 3, pp. 393–432, 2016.
- [32] J. Santana, R. Hoover, and M. Vengadasubbu, "Investor commitment to serial entrepreneurs: a multilayer network analysis," *Social Networks*, vol. 48, pp. 256–269, 2017.
- [33] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.
- [34] N. E. Friedkin, "Theoretical foundations for centrality measures," *American Journal of Sociology*, vol. 96, no. 6, pp. 1478–1504, 1991.

- [35] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [36] K.-I. Goh, E. Oh, H. Jeong, B. Kahng, and D. Kim, “Classification of scale-free networks,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 20, pp. 12583–12588, 2002.
- [37] K. I. Goh, B. Kahng, and D. Kim, “Universal behavior of load distribution in scale-free networks,” *Physical Review Letters*, vol. 87, no. 27, Article ID 278701, 2001.
- [38] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, “Vital nodes identification in complex networks,” *Physics Reports*, vol. 650, pp. 1–63, 2016.
- [39] S. J. Chang, “Venture capital financing, strategic alliances, and the initial public offerings of Internet startups,” *Journal of Business Venturing*, vol. 19, no. 5, pp. 721–741, 2004.
- [40] J. Lerner, A. Schoar, S. Sokolinski, and K. Wilson, “The globalization of angel investments: evidence across countries,” *Journal of Financial Economics*, vol. 127, no. 1, pp. 1–20, 2018.
- [41] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic Regression*, Springer, New York, NY, USA, 2002.
- [42] D. A. Kenny, *Statistics for the Social and Behavioral Sciences*, Brown, Providence, RI, USA, 1987.