Research internship report

Network Science applied to the VC industry

In this research internship my work was mainly centered on two different sub-projects: the first one focused on the analysis and filtering of tweets and the second one, which was the main one, focused on the study of the structure and interactions between companies and investors in the Venture capital ecosystem, through the use of Networks.

Tweets filtering

In this first project the objective was to extrapolate some top VC firms' tweets and through NLP, in particular a process of lemmatization and context extrapolation, group them into four subcategories: balance sheet, csuite strategy, innovation and employment. Then, the following scores were computed: finBERT, roBERTa, Vader compound score, total Vader, non-firm Vader. Also the total likes, non-firm likes and firm likes were taken into consideration. Finally, for every different dataset, the mean of the scores and the sum of the likes were computed considering the period of time equal to three years before every tweet date belonging to a specific firm. My role was also the analysis, synthesis, arrangement and manipulation of datasets with a huge number of entries.

Venture capital networks

The main part of my internship was to contribute in the examination of the role of networks in shaping a company's chances of raising series A funding, based on the investors who backed it in its earliest stages. We did so by considering ways in which some investors can, more than others, play a role in the attraction of capital towards the startups they back. In particular, one of the main questions we tried to answer was: to what extent (if at all) do high attraction investors participating in seed rounds influence the startup's chances of raising series A, and with which results?

The methodology used followed three main stages:

1. Creation of a network of investors and startups sensible to temporal dynamics
2. Ideation and formulation of specific scores able to capture the attraction capacity of the investors and help understand leader-follower dynamics
3. Projection of such scores on the companies which are about to raise series A

The investments dynamics that occur between early investors and companies have a structure that can be naturally described my means of a bipartite network. The vertices of the network are of two types: early investors *i* and companies *c*. For every investor *i* and company *c* we draw an oriented edge from i to c, if investor *i* participates in a funding round of company *c*. In this case, nodes *i* feature exclusively outgoing edges, called out-degrees *ki*;, while nodes *c* display only incoming edges, called in-degrees *kc*.
In the bipartite network we built, each company is open to investments twice (round 1 and round 2), while investors can participate multiple times according to their endowment. In studying the passage from pre-series A to series A, round 1 comprises all the pre-series A rounds (angel, pre-seed, seed), while round 2 corresponds to series A. Similarly, when the network is re-created to

analyse the advancement from series A to series B, round 1 displays the seeding round plus series A while round 2 represents series B.
The network grows in a discrete fashion from 2010 to 2021. Funding interactions and new nodes are inserted at the beginning of every month, creating a snapshot sequence of graphs.
The discrete time dynamic allows to represent complex temporal patters not captured by static networks. Furthermore, it explicitly ensembles complex temporal patterns that are no captured by static networks, and has the potential to significantly enhance our understanding of the dynamics underlying early stage funding.

To analyse the role of investors and specifically their ability to catalyze investments on the companies they back, we developed a novel network measure. We based the creation of the innovative index on the topological overlap matrix, a similarity matrix built on the overlap of neighbors between two nodes. Indeed, the matrix reflects how many neighbour nodes are connected to both the vertices and weights them by the minimum number of degrees between the vertices. It measures pairwise similarities of nodes: the greater the similarity of two objects, the greater the value of the measure.
In its original version, the matrix is static, refers to a univariate network, and to pair-wise relations. Initially, we built the matrix in a bivariate setting, having the investors nodes *i* in the axes and considering companies nodes *c* as the connecting neighbours.
Consequently, we added a temporal dimension, shifting the values of one axis to the following moment in time. Lastly, the communal aspect was taken into account, enabling the computation of a score which captures a one-to-many leader-followers dynamic, defined as the influence that the single investor *i* has on the community of investors j.
Applied to our case, the *Ti* index considers each investor at round 2, *j*, which is funding the company c in which *i* invested at round 1 and scale its investment by the minimum number between investments in first rounds made by *i* and investments at second round made by *j*. Consequently, the ratios attributed to all investors *j* related to investor *i* are averaged, resulting in the *Ti* score. The *Ti* score is to be intended as an intermediary step between the topological overlap matrix and the funding attraction index *Mi*, as it lacks the economic significance of the latter, but it enriches with the bipartite, the temporal, and the communal dimension the former.

Based on it, we introduce the funding attraction index (*Mi*) that quantifies the number of times that the investments of an early investor at round 1 have been replicated by other early investors at round 2. Therefore, an investor *i* participating in round 1 is assigned a score based on how many investors j participating at round 2 are funding the company *c* in which *i* invested at round 1, scaled by the number of investments made by *i* at round 1. The result is then normalized to assure that the funding attraction index lies in the range [0, 1]. By itself, measure *Mi*, quantifies the capability of an investor *i* to attract, at a subsequent round, further investors in the company they invested in. It is defined as an investment multiple assigned to the investor and perfectly captures the one-to-many leader-followers dynamic which it meant to ensemble.

In order to capture the one-to-one leader-followers dynamic, a similar score (*Lij*) was computed: it encapsulate the same reason and scope of the *Mi* score, but with the difference that it is not scaled by the number of investments made by *i* at round 1, but by the number of investments made by *j* at round 2, and it is pairwise: indeed, it allowed us to create a matrix of pair-wise comparison between leader and follower, which would tell us how much of what the follower did at round 2 was influenced by what the leader did.

This matrix also allowed us to select any pair of leader-follower and see the pair-wise evolution of the score over time, and also try to understand whether being a Midas could also be a factor of influence.

To test whether companies backed by investors with a high funding attraction factor are more likely to raise series A, we explored the interaction between differently ranked investors and companies funding attraction. Through different OLS regression models we computed the likelihood of raising series A funding and the actual amount of money raised in series A.
The empirical assignment is straightforward: to test the hypotheses that

1. a company getting funding from high attraction investors in the seed rounds is more likely to raise a series A round, while the mechanism does not hold when advancing from series A to series B and,
2. it raises more funds in the follow-on round.

in both univariate and multivariate settings.

In conclusion, the project was aimed at testing, through the lenses of network theory, that some investors are more efficient than others in making the companies they back obtain series A funding, and examined their impact on the amount of capital attracted during the follow-on round.
Indeed, early investors with a high funding attraction score favour young companies in attracting further funding and lead them to more conspicuous rounds.
The results help deepen our understanding of the mechanism ruling how a company elevates from the uncertain ground of seed financing by picking the angels and venture capitalists who can better support them in their seek for the limelight. Being known and proceeding in business life is essential for any company determined to gain chances to succeed.
The innovative transitive fitness model allows us to make the company the protagonist of the ultimate model, while searching for answers looking at investors.