TechRank

Anita Mezzetti, Loïc Maréchal, Dimitri Percia David, Thomas Maillart, and Alain Mermoud

Anita Mezzetti
is a financial engineer at EPFL and a quant at Credit Suisse in Zurich, Switzerland.
anitamezzetti97@gmail.com

Loïc Maréchal
is a researcher at the Cyber-Defence Campus, Armasuisse Science and Technology, and HEC Lausanne in Lausanne, Switzerland.
loic.marechal@unil.ch

Dimitri Percia David
is an assistant professor of data science and econometrics at the University of Applied Sciences of Western Switzerland in Valais, Switzerland.
dimitri.perciadavid@hevs.ch

Thomas Maillart
is a senior lecturer at the Geneva School of Economics and Management in Geneva, Switzerland.
thomas.maillart@unige.ch

Alain Mermoud
is the head of technology monitoring and forecasting at the Cyber-Defence Campus, Armasuisse Science and Technology, in Lausanne, Switzerland.
mermouda@ethz.ch

KEY FINDINGS

- This article introduces a recursive algorithm based on a bipartite graph linking companies and technologies.
- The authors' method overcomes the typical caveats of asset pricing in the context of private equity, where cash flows are not observable.
- The authors' method is flexible enough to allow investors to plug their preferences directly into the model.

ABSTRACT

This article introduces TechRank, a recursive algorithm based on a bipartite graph with weighted nodes that the authors developed to link companies and technologies based on the reflection method. They allow the algorithm to incorporate exogenous variables that reflect an investor's preferences and calibrate the algorithm in the cybersecurity sector. First, their results help estimate each entity's influence and explain companies' and technologies' ranking. Second, the results provide investors with an optimal quantitative ranking of technologies and thus help them design their optimal portfolio. The authors propose this static method as an alternative to traditional portfolio management and, in the case of private equity investments, as a new way to optimize portfolios of assets for which cash flows are not observable.

This work investigates the innovation structure and the dynamics underlying the life cycle of technologies. We fill two research gaps—the first concerns identifying future benefits and risks of emerging technologies for society. The second regards the valuation of early-stage companies and optimal investment decisions. To fill these gaps, we introduce the TechRank algorithm. Our methodology assigns a score to each entity—that is, technologies and firms—based on their contribution to the technological ecosystem. We expect this method to help stakeholders form optimal investment, procurement, and technology-monitoring decisions.

We calibrate our model in the cybersecurity sector, although TechRank could apply to any sector. The cybersecurity technological landscape represents a challenge for this calibration, given the important share of startups and the fast innovations it yields (Gordon et al. 2018). Moreover, the important number of cyberattacks and the increasing costs they incur have boosted cybersecurity investments.[1]

[1] Erin Woo, "As Cyberattacks Surge, Security Start-Ups Reap the Rewards," *The New York Times*, July 26, 2021. Tipranks, "Microsoft Securing its Position with Cybersecurity Investments," *Yahoo Finance*, July 20, 2021.

According to Bloomberg, "the global cybersecurity market size is expected to reach USD 326.4 billion by 2027, registering a compound annual growth rate of 10.0% from 2020 to 2027."[2] An additional justification for choosing this sector is that private equity markets are extremely idiosyncratic, thus leading private equity investors to become specialists in their fields and often invest in firms of a single sector.

To develop the TechRank algorithm, we first model and map the ecosystem of companies and technologies from the Crunchbase dataset using a bipartite network (Crunchbase, Inc. 2022). The bipartite network structure accurately describes this complex and heterogeneous system. We evaluate the relative influence of the network nodes in the ecosystem by adapting a recursive algorithm that estimates network centrality.

This methodology should help decision-makers and investors assess entities' influence in the cybersecurity ecosystem, thereby reducing investment uncertainties. Around 90% of startups fail, and in 42% of the cases, the failure stems from incorrect evaluation of the market demand. The second reason (29%) is that they run out of funding and personal money.[3] Christensen (1997) highlights that well-managed companies also break down because they overinvest in new technologies (Christensen 1997). Thus, selecting the right technologies to invest in aligns with the optimal investment strategy.

Our research takes inspiration from Google's PageRank algorithm, which ranks web pages according to readers' interests (Page et al. 1999). We use a similar approach with bipartite networks to assign a score to companies and technologies, expanding the methodology by adding the impact of external factors. Our method is flexible and permits the incorporation of investors' preferences, such as location or previous funding rounds. TechRank lets the investor select entities' features that reflect their interests. The algorithm uses their choices as input, which tweaks the entities' scores to reflect them. This enables investors to select a personalized portfolio strategy using a quantitative methodology. Evaluating companies and new technologies largely depends on investors' personal choices, which may lead to misreading market demand. This work aims to lead to more systematic decision-making for investors.

The remainder of this article proceeds as follows. The next section presents the literature review and hypotheses. The third section details the data and the methodology. The fourth section presents the results. The final section concludes.

LITERATURE REVIEW AND HYPOTHESES DEVELOPMENT

Centrality Measures

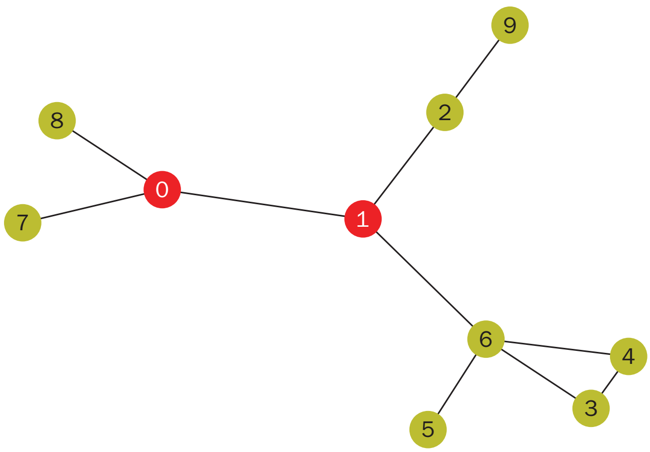
Exhibit 1 illustrates the difference between central and peripheral nodes in a graph. In network analysis, a centrality measure estimates the importance of nodes through ranking. The most simple centrality estimate is the "degree," which counts the number of neighbors of a node. One of its drawbacks is that it does not show which one stands in the center of the network. Two nodes may share the same degree while being more or less peripheral. Thus, the degree is a local centrality measure, which does not capture the influence across nodes within the graph.

Another important centrality measure is "closeness," which measures how long information spreads from one node to the next. Specifically, closeness is defined as the reciprocal of "farness"—that is, the sum of distances between one node and

[2] "Global Cybersecurity Market Could Exceed $320 Billion in Revenues by 2027," *Bloomberg,* July 29, 2020.

[3] Jack Stemward, "The Ultimate List of Startup Statistics for 2021," *Findstack*.

EXHIBIT 1
Central and Peripheral Nodes



NOTE: This exhibit depicts the difference between central (red) and peripheral (acid green) nodes in a graph.

all other nodes. The "betweenness centrality" of a node measures how often a node stands in the shortest path between a pair of other nodes (e.g., see Bavelas 1948; Saxena and Iyengar 2020; Freeman 1978).

Another strand of research focuses on the top-K shortest path identification in a complex network, a topic less tackled by the literature than centrality. To rank nodes, one must compute the centrality of all nodes and compare them to extract the rank, which is not always feasible due to the network size. To overcome this problem, Saxena and Iyengar (2017) attempt to estimate the global centrality of a node without analyzing the whole network. Similarly, Bavelas (1948) develops a structural centrality measure in the context of social graphs. Other centrality concepts include the eigenvector, Katz, or PageRank centralities (Bonacich 1972; Page et al. 1999; Katz 1953). Finally, Freeman (1978) creates a formal mathematical framework for centrality, including degree, closeness, and betweenness, and advocates for combining different kinds of centrality measures.

PageRank

Page et al. (1999) developed the PageRank algorithm. Its primary goal is to rank web pages objectively, a challenge with the fast-growing web. PageRank assigns a score to each web page based on its relations with other web pages in the graph. Other fields have benefited from PageRank providing modifications and improvements. Xing and Ghorbani (2004) extended the algorithm and proposed the weighted PageRank (WPR). This algorithm assigns larger rank values to more important pages instead of dividing the rank evenly among its outlink pages.[4] Each outlink page gets a value proportional to its popularity, considering the weights of the links. One caveat of PageRank and its variants is that they do not consider n-partite structures; yet, web pages can all be linked to one another. Bipartite networks address this issue and capture this complexity, among other structures.

Bipartite Networks

Networks are a fundamental tool to capture the relations between entities. Graphs (G) are composed of vertices (V) and edges (E), and we denote $G = (V, E)$. We build links and mathematically analyze many properties of the whole system and singular entities. To graphically represent a real system, we synthesize its information into a simple graph framework. This simplification generates an information loss in the modeling process. Simple network structures might discard important information about the structure and function of the original system (Kurant and Thiran 2006). Consequently, the failure of a tiny fraction of nodes in one network may lead to the complete fragmentation of a system (Buldyrev et al. 2010). To solve the problem, extensions to the simple structure $G = (V, E)$ are added, as well as yield graphs with more powerful features. For instance, in the case of vertices connected by relationships of different kinds, Battiston, Nicosia, and Latora (2014) advocate working with

[4] Given a web page W, an inlink of W is a link to another web page that includes a link pointing to W. An outlink of W is a link appearing in W, which points to another web page.

multiplex networks—that is, networks where each node appears in a set of different layers, and each layer describes all the edges of a given type. When it is possible to distinguish the nature of the edges, multiplex networks are an effective approach, which starts by embedding the edges in different layers according to their type. However, even if we have two kinds of nodes, the nature of the edges is unique. Therefore, a more suitable approach is bipartite networks. Bipartite networks are, for instance, a good way to describe the technological and business landscape. Exhibit 2 depicts two sets of interconnected nodes, companies, and technologies, which do not present edges within the same set.

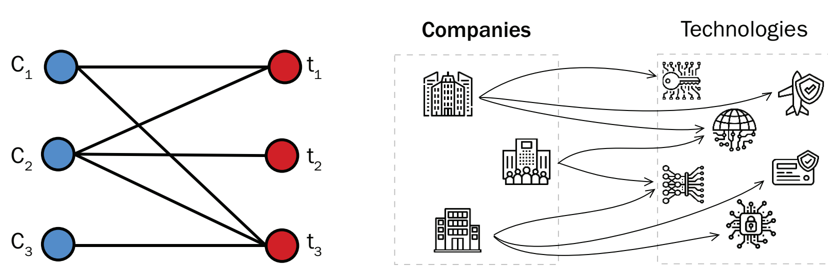
There are multiple adaptations of the PageRank algorithm to bipartite structures. In particular, Benzi, Estrada, and Klymko (2013); Donato et al. (2004); and Tu et al. (2018) extend the PageRank algorithm to multiplex networks. They assume that only some graph clusters are multiplex networks and extend the PageRank algorithm only to analyze the subgraph centrality. Bipartite networks transform directed into undirected networks with twice the number of vertices.

Klein, Maillart, and Chuang (2015) extend PageRank in the Wikipedia editors and articles context. The application of this algorithm to the case of interactions between companies and technologies is straightforward. A major benefit of this approach is that it starts from an unweighted graph, linking authors and articles. They develop a recursive algorithm in which the two entities contribute to the quality (for articles) or the expertise (for authors) of each other. They develop a bipartite random walker by building the adjacency matrix $M_{e,a}$ that takes the value one if editor e has edited article a and 0 otherwise, which tracks all the editors' contributions. They obtain $M_{e,a} \in \mathbb{R}^{n_e, n_a}$, where n_e and n_a are the number of editors and articles. They sort editors by the number of articles' contributions and assign a contribution (quality) value to each editor (article) based on their degree. The expertise w_e^0 (quality w_a^0) is given by the number of editors (articles) that have worked on articles (have received modifications).

The second part of the algorithm follows a Markov process in its iterations. The step w^n ($w^n = w^n(\alpha, \beta)$) depends only on the information available at w^{n-1}. At each step, the algorithm incorporates information about the expertise of editors and the quality of articles within the bipartite network structure. The process is a random walker with jumps, whose transition probability is zero in the case of $M_{e,a} = 0$. Next, the authors define two variables for the transition probability, $G_{e,a}(\beta)$ and $G_{e,a}(\alpha)$. $G_{e,a}(\beta)$ represents the probability of jumping from article a to editor e, and $G_{a,e}(\alpha)$ represents the probability of jumping from an editor to an article. Both parameters depend on initial conditions. The optimal parameters are selected through a grid search that maximizes the Spearman rank correlation between the rank given by the model and ground-truth metrics obtained independently. Finally, Klein et al. (2015)

EXHIBIT 2

Bipartite Structure of Companies and Technologies



NOTES: The left panel depicts a typical bipartite structure. The right panel illustrates this structure with companies (layer 1) and technologies (layer 2). Image: Flaticon.com.

observe a “less-is-more” situation, because too many editors working on an article is detrimental to its quality. Studying different categories of Wikipedia articles, they find α to remain constant, while β varies significantly across categories.

Estimating the global rank of a node starting from local information and centrality measures is still an open research question in many sectors (Saxena and Iyengar 2020). This method is vastly overlooked in many subfields of finance, with the noticeable exception of Konstantinov (2022), who uses graph theory to estimate risk flows across clusters of hedge funds. Instead, no research uses this approach for investment decisions and portfolio optimization in the private equity sector, to the best of our knowledge. Yet, this approach could help overcome the limitations of standard financial models in private equity, where the network structure is easily obtainable but the cash flow process is not.

Private Equity Valuation

Private firms are not required to publicly disclose their financial statements, which makes it difficult to measure their past performance and estimate their expected returns without insider information. Moreover, because they are not listed on exchanges, we do not observe the expectations of market participants. Thus, standard asset pricing methods fall short in this context. Similarly, private equity analysts must rely on insider or private information to value private firms. These valuations generally occur around a financing round and have to take into account the capital dilution to compute realized returns on the firm (Gornall and Strebulaev 2020).

A first generic approach, which attempts to overcome these limitations and estimate venture capital’s expected returns and risk, uses private equity or venture capital fund–level observations. For instance, Kaplan and Schoar (2005) study the performance of private equity partnerships and find that, after fees, the average fund returns are on par with those of the S&P 500 Index, albeit with significant heterogeneity across funds. They additionally document a return persistence in funds returns. In contrast, in the venture capital sector, Harris, Jenkinson, and Kaplan (2014) find that the outperformance of venture capital funds exceeds that of the S&P 500 Index by about 3% per year. They also uncover that this performance is time-varying, with significant positive excess returns in the 1990s but negative returns in the 2000s. To overcome the challenges of data availability and databases’ heterogeneity that yields different results, Harris, Jenkinson, and Kaplan (2016) use cash flow data of 300 institutional investors involved in about 1,800 venture capital funds in the United States. They confirm the findings of Harris et al. (2014) and find that fund returns exceed those of public markets for all vintage years but one. They also confirm the average excess returns on the market of about 3% and the cyclicality of these returns. They additionally link the time variation in returns to capital flows, documenting that returns are larger (smaller) for funds that started when there were small (large) capital inflows to the sector.

Brown et al. (2021) extend the cyclicality analysis and offer the same conclusions. In contrast, they do not find that an investable strategy that invests in the time of small inflows and divests in the time of large inflows would generate additional performance. Braun, Jenkinson, and Stoff (2017) and Harris, Jenkinson, Kaplan, and Stucke (2023) study in depth the persistence properties of returns in private equity, already documented by Kaplan and Schoar (2005). On the one hand, Braun et al. (2017) find that the persistence of funds’ performance has declined along with the sector’s maturity and increased competitiveness, leading to insignificant autocorrelation in the performance process as for other asset classes. On the other hand, Harris et al. (2023) find that, on average, performance is persistent for funds raised by the same general partner. However, they confirm that their performance does not persist regarding buyout funds. Other private equity and venture capital studies include the return manipulation by managers around fundraising (Brown, Gredil, and Kaplan 2019), the diversification property of private equity (Brown, Crouch, Ghent, Harris, Hochberg, Jenkinson, Kaplan, Maxwell, and Robinson 2022), the funds of funds in private equity (Harris, Jenkinson, Kaplan, and Stucke 2018), or the liquidity property of private equity cash flows (Robinson and Sensoy 2016).

A second approach to overcome the estimations difficulty is to use the information from firm observations at their intrinsic level. Cochrane (2005). He uses a maximum-likelihood estimation method to obtain these values at the market and sector levels, such as healthcare, biotechnology, technology companies, and retail services. He finds a mean arithmetic return of 59%, an alpha of 32%, a beta of 1.9, and a volatility of 86% (equivalent to a 4.7% daily volatility). Given that the distribution of returns is heavily positively skewed in venture capital, he adopts a logarithmic model that also accounts for the inherent selection bias of this asset class.[5] Ewens (2009) updates this method on returns computed from one financing round to the next. He adopts a three-regime mixture model (failure, medium returns, and "home runs"). He also corrects for the selection bias and obtains an alpha of 27% and a beta of 2.4. He finds that 60% of all venture capital investments have a negative log return. Altogether the results are similar. Venture capital investments exhibit positive alpha, large beta, and high volatility. Other attempts to evaluate the market parameters of the venture capital asset class include Korteweg and Nagel (2016) and Moskowitz and Vissing-Jørgensen (2002), with results in line with the previously mentioned studies.

Another strand of research attempts to index and benchmark the private equity market. Peng (2001) builds a venture capital index from 1987 to 1999 from about 13,000 financing rounds targeting more than 5,600 firms. He addresses the problems of missing data, censored data, and sample selection by using a reweighting procedure and method of moment regressions. From the index perspective, the results are qualitatively the same—that is, high and volatile returns to venture capital (average return of 55.18% per year). He finds his index to display much higher volatility than the S&P 500 and NASDAQ indices and high exposure to these indices (betas of 2.4 and 4.7, respectively). Other venture capital indices construction includes Hwang, Quigley, and Woodward (2005); Schmidt (2006); Cumming, Haß, and Schweizer (2013), who all obtain results on par with the previously mentioned studies.

One limitation of the previous studies is that they estimate these parameters only at an aggregate level. An investor could form her investment decisions and portfolio choices by segregating among sectors but not obtaining the actual firms' parameters. One exception is Moon and Schwartz (2000), who provide an approach based on the real-options theory to price individual firms. However, this method requires the observations of cash flows, and they provide only one calibration example with Amazon. Another alternative consists of analyzing the potential markets of early-stage firms but is restricted to technological firms (Andries, Clarysse, and Costa 2021). The last approach uses funding rounds or corporate-level information to predict the success of early-stage firms (e.g., see Cumming 2006; Liu 2021). However, these methods do not leverage the whole set of information available. Thus, there remains a caveat in methodology to help investors form optimal decisions using all the information available. Given the recent venture capital boom, Zhong et al. (2018) advocate for the use of quantitative methodologies of screening and evaluation. However, there is a clear research gap in methodologies enabling one to value early-stage companies and to form optimal portfolios. These methodologies either enable one to value only a sector instead of a specific company or require the use of cash flows that are unobservable.

[5] Most of venture capital data is private, and available data are more often related to successful firms than underperforming ones.

Nonfinancial features and relations between companies, technologies, founders, and investors are instead numerous and easily observable (e.g., see Dalle, den Besten, and Menon 2017; Smith, Smith, and Shaw 2017; Baron and Markman 2003; Hoang and Antoncic 2003). We thus formulate our hypotheses as follows:

H_1: *Using a bipartite network structure allows an algorithm to rank companies based on their links with technologies*.
H_2: *This algorithm and its ranking may be improved and tilted toward investors' preferences*.
H_3: *The ranking performance is independent of the sector considered*.

DATA AND METHODS

Data

We use Crunchbase data.[6] Crunchbase is a commercial database that provides access to financial and managerial data on private and public companies globally. It was created in 2007 by TechCrunch, a source of information about startup activities and their financing within and across countries. This database has mainly been adopted by both academics and industry practitioners (den Besten 2021; Fisch and Block 2021). It is also used by international organizations such as the Organisation for Economic Co-operation and Development (OECD) (Dalle et al. 2017).

Crunchbase is made of data collected with a multifaceted approach that combines crowd-sourcing (through venture programs or direct contributions), machine learning, in-house processing, and aggregation of third-party providers' data. Crunchbase updates and revises data daily and organizes it into several entities such as organizations, people, events, acquisitions, or IPOs. The primary focus of Crunchbase is the technology industry, although it also includes data on other sectors.

Data can be accessed in two ways: using an API or downloading a comma-separated values (CSV) file directly from the Crunchbase website. Data are split into several databases depending on their type. We provide a nonexhaustive list in Exhibit 3.[7]

We first analyze the Crunchbase dataset dedicated to investors. With a total of 185,784 investors divided into 78,001 (41.98%) organizations and 107,783 (58.02%) persons, there are more investors than target companies. Exhibit 4 shows that most investors are pure (87.11%). Some organizations are both investees and investors (12.65%). The remainder of the sample is typically universities. Crunchbase ranks

EXHIBIT 3

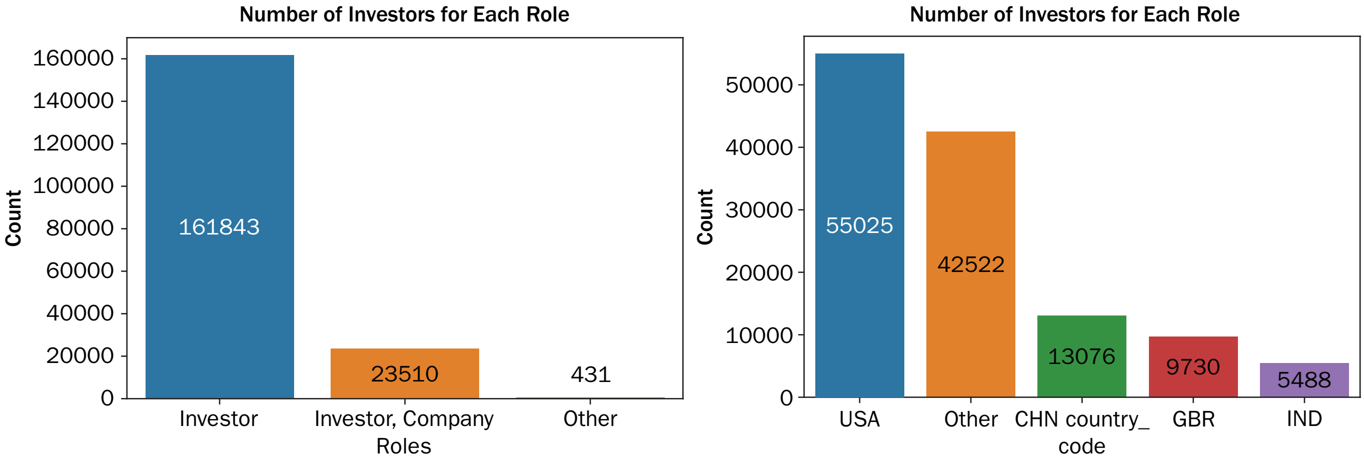
Crunchbase Files Description

Field Name	Description
organizations	Organization profiles
organization_desc	Long descriptions of organization profiles
acquisitions	List of all acquisitions
org_parents	Map between parent organizations and subsidiaries
IPOs	Detail for each IPO
people	People profiles
people_desc	Long descriptions for people's profiles
degrees	Details of people's educational background
jobs	List of all jobs and advisory roles
investors	Active investors (organizations and people)
investments	All investments
investment_partners	Partners are responsible for their firm's investments
funds	Details of investments funds
funding_rounds	Details for each funding round in the dataset
events	Event details
event_appearances	Event participation details

NOTE: This exhibit reports the main fields available from Crunchbase.

[6] Crunchbase website: https://www.crunchbase.com/.
[7] Crunchbase daily CSV files were exported from https://data.crunchbase.com/docs/daily-csv-export. Data were downloaded on April 28, 2021.

EXHIBIT 4
Summary Statistics of Investors in Crunchbase Data



NOTES: The left panel depicts the distribution of investors according to their types. The right panel depicts the number of investors per country.

the top 1,000 investors through its proprietary algorithm. Exhibit 4 indicates that most investors are in the United States (29.62%). In particular, there is a wide gap between the first and second countries, China, where 7.04% of investors are located.

Methodology

Adaptation of Klein et al. (2015). In this research, we use a bipartite network that describes the relations among companies and the technologies they are involved in. Exhibit 2 describes the typical bipartite network structure. This structure benefits from advances in network theory, Markov chains, and machine learning. We adapt the recursive algorithm with the method of Hidalgo, Hausmann, and Dasgupta (2009). We expect the resulting rank to incorporate the positive influence of well-established companies on technologies and, at the same time, the positive impact of new companies that explore new fields. We build the adjacency matrix $M^{CT}_{c,t} \in \mathbb{R}^{n^c,n^t}$, which takes a value of 1 if a company c works on a technology t and 0 otherwise. n^c and n^t represent the number of companies and technologies. We assume that well-established companies have more means to diversify their expertise and, therefore, that an entity has a relatively high number of neighbors (Canito et al. 2018; Gold, Malhotra, and Segars 2001). Thus, we initialize the algorithm with the degree—that is, counting the neighbors of each entity,

$$\begin{cases} w_c^0 = \sum_{t=1}^{n^t} M^{CT}_{c,t} = k_c \\ w_t^0 = \sum_{c=1}^{n^c} M^{CT}_{c,t} = k_t \end{cases} \tag{1}$$

The algorithm is a "random walker" that incorporates information about company expertise and technology relevance at each step. The transition probabilities, $G_{c,t}$ and $G_{t,c}$, describe the extent to which the entities' weights change over the iterations. If the relation between c and t increases (decreases) the value, the entity weight increases (decreases) in proportion with the transition probabilities. We define $G_{c,t}$ and $G_{t,c}$,

$$\begin{cases} G_{c,t}(\beta) &= \dfrac{M^{CT}_{c,t} k_c^{-\beta}}{\sum_{c'=1}^{n^c} M^{CT}_{c',t} k_c'^{-\beta}} \\ G_{t,c}(\alpha) &= \dfrac{M^{CT}_{c,t} k_t^{-\alpha}}{\sum_{t'=1}^{n^t} M^{CT}_{c,t'} k_t'^{-\alpha}} \end{cases} \tag{2}$$

where α and β inform how coordination generates value. Next, we define the recursive step,

$$\begin{cases} w_c^{n+1} &= \sum_{t=1}^{n^t} G_{c,t}(\beta) w_t^n \\ w_t^{n+1} &= \sum_{c=1}^{n^c} G_{t,c}(\alpha) w_c^n \end{cases} \tag{3}$$

As in PageRank, the recursion ends when the algorithm converges. Our algorithm allows for considering the market complexity and feedback loops (investments' impact on companies and technologies). After adding exogenous factors, we discuss this feature and the optimization of α and β.

Inclusion of exogenous factors. We include exogenous factors as the ground truth in the parameters' calibration step. This allows for keeping the algorithm tractable while letting it capture the technological structure. We use this ground truth to compute the Spearman correlation—ρ_c for companies and ρ_t for technologies. Because ρ_c and ρ_t depend on α and β (see Equation 2), we find the parameters that maximize these correlations,

$$\begin{cases} (\alpha^*, \beta^*) = \arg\max_{\alpha,\beta} \rho_c(\alpha,\beta) \\ (\alpha^*, \beta^*) = \arg\max_{\alpha,\beta} \rho_t(\alpha,\beta) \end{cases} \tag{4}$$

and we solve this optimization problem with a grid search. Equation 4 shows that parameters depend on companies and technologies. This dependence enables the creation of the structure of the bipartite graph. To obtain the correlation between the TechRank score, which assigns a weight w_c (w_t) to each company (technology), and the ground truth evaluation, which assigns $\hat{w}_c$ ($\hat{w}_t$) to each company (technology), we normalize both TechRank results and the exogenous measure in the same range [0, 1].

Investors use the entities' features to select companies and the investment amount they want to allocate. We suppose that an investor has $n^{(C)}$ features to pick from, denoted as $f_1^{(C)},\ldots,f_{n^{(C)}}^{(C)}$, where C (T) represents the association with the companies (technologies). Each feature $f_i^{(C)}$ is associated with a percentage of interest $p_i^{(C)}$ and $\sum_{i=0}^{n^{(C)}} p_i^{(C)} = 1$. For instance, if a company's feature is the previous investment amount and geographical proximity to the investor, $n^{(C)} = 2$. An investor may then decide to be interested at 80% in the first feature and 20% in the second, by selecting $p_1^{(C)} = 0.8$ and $p_2^{(C)} = 0.2$. A feature may also push investors back, so we multiply it by –1. We define all notations in Exhibit 5.

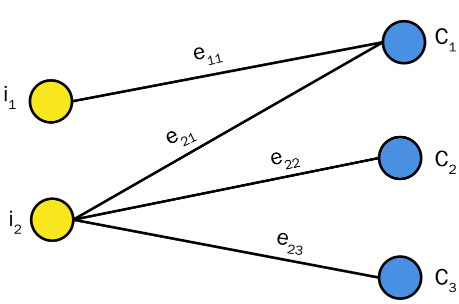
We convert quantitative and qualitative properties into a number $f_i^{(C)} \in [0,1]$. Next, we create all the factors $f^{(C)} = f_1^{(C)},\ldots,f_{n^{(C)}}^{(C)}$ for the estimations. We provide the full equations and their descriptions in the appendix, for the estimation of the exogenous evaluation $\hat{w}_c$ in Equation 5 and for the system for the exogenous evaluation for companies and technologies $\hat{w}_c$ and $\hat{w}_t$ in Equation 6. To select the features, we use Crunchbase data about companies and investors (see Exhibit 3).

EXHIBIT 5
Variable Definitions

Variable	∈	Description
$n^{(C)}$	$\mathbb{N}$	Number of external features available for companies.
n^{c}	$\mathbb{N}$	Number of companies.
$n^{(T)}$	$\mathbb{N}$	Number of external features available for technologies.
n^{t}	$\mathbb{N}$	Number of technologies.
$p_i^{(C)}$	[0, 1]	Percentage of interest in the company preference number *i*.
$p_j^{(T)}$	[0, 1]	Percentage of interest in the technology preference number *j*.
$f_i^{(C)}$	$\mathbb{R}^{n^c}$	Vector of factors associated with the company preference number *i*.
$f_j^{(T)}$	$\mathbb{R}^{n^t}$	Vector of factors associated with the technology preference number *j*.
n^{i}	$\mathbb{N}$	Number of investors.
M^{CT}	$\mathbb{R}^{n^c, n^t}$	Adjacency matrix of the C-T bipartite network
M^{IC}	$\mathbb{R}^{n^i, n^c}$	Adjacency matrix of the I-C bipartite network
$\gamma_t^{i,c}$	$\mathbb{R}$	Amount in funding round between *c* and *i* at time *t*
e^{IC}	$\mathbb{R}^{n^i, n^c}$	Total amount of investment between each investor to each company
e^{C}	$\mathbb{R}^{n^c}$	Total amount of investments toward each company
e^{T}	$\mathbb{R}^{n^t}$	Total amount of investments toward each technology
e_{max}^{C}	$\mathbb{R}$	Maximum amount of total investments among all the companies
e_{max}^{T}	$\mathbb{R}$	Maximum amount of total investments among all the technologies
f_c^{C}	[0, 1]	Factor related to previous investments into the company number *c*
f_t^{T}	[0, 1]	Factor related to previous investments into the technology number *t*

NOTE: This exhibit presents the variable definitions used throughout the article.

EXHIBIT 6
Investors–Companies Bipartite Network



NOTE: This exhibit depicts a stylized bipartite network between investors and companies.

Previous Investments

We assume that previous investment is an essential factor in evaluating companies. Investors may be willing to invest in companies that have already received capital or look for higher returns, targeting newer firms.

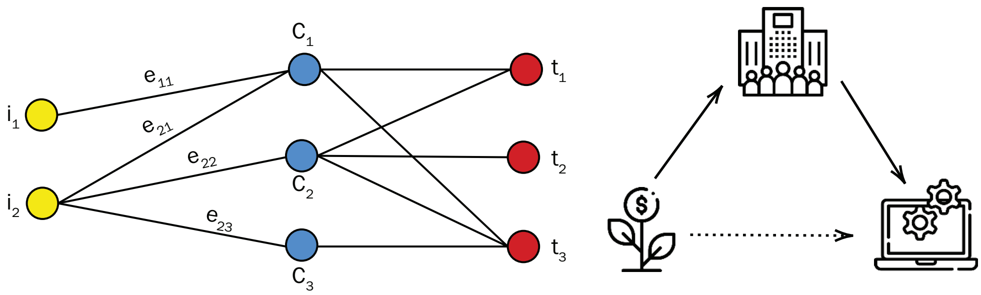
To compute this factor, we use the Crunchbase field "funding_rounds," which reports the amount of all funding rounds from an investor *i* to a company *c*. We capture this structure with another bipartite network describing the links between investors (I) and companies (C). In this case, we weigh the edges by the sum of all previous investments from investor *i* to company *c* until the current period ($\mathcal{T}$) and compute the adjacency matrix M^{IC}. We define the amount of a single investment from *i* to *c* at time *t* by $\gamma_t^{i,c}$. The weight of the edge $i - c$ is given by $e_{i,c} = \sum_{t=0}^{\mathcal{T}} \gamma_t^{i,c}$ (see Exhibit 5). We then sum the contribution of all investors to find the attribute $f_c^C \in [0,1]$ for a company c.[8] Next, we normalize by dividing each investment by the largest investment of our sample.

Exhibit 6 depicts the investment structure as an example. We consider two investors i_1 and i_2 and three companies C_1, C_2, and C_3. We compute the maximum e_{max} as $\max\{e_{11} + e_{21}, e_{22}, e_{23}\}$. In the appendix, we report the features related to the investments for each company in Equation 7 and its generalization in Equation 8.

[8] Note that here f_c^C represents the factor related to a company.

EXHIBIT 7

Tripartite Structure of Investors, Companies, and Technologies



NOTES: The left panel depicts a typical tripartite structure. The right panel illustrates this structure with investments (layer 1), companies (layer 2), and technologies (layer 3). Image: Flaticon.com.

We link the two bipartite structures, investment–companies, and companies–technologies, to obtain an I-C-T tripartite structure depicted in Exhibit 7. This structure allows assigning some features to technologies from companies (direct link) or investors (indirect link). Thus, we can find the previous investment in technology through companies' funding rounds. We report the generalized equations for the previous investment's factor for technology, Equation 9, and the corresponding algorithm in the appendix.

Location

The second feature we consider is the distance between investors' and companies' locations. We retrieve the addresses of firms and investors from Crunchbase (*c_address*) and map them to geographic coordinates. We compute the Haversine approximation to measure the distance. We detail the Haversine approximation in the appendix. Investors may prefer short-distance investments or places with high potential. If they face some investment restrictions, we filter the companies based on the criteria before applying the algorithm. Otherwise, we add a distance factor to the algorithm.

We use the Haversine distance h to obtain a factor $f_c^{(C)} \in [0,1]$ for each company. We consider the distance $h_{i,c}$ between the company c and investor i. We assume that the factor is the proximity, so that $f_c^{(C)}$ tends to 1 as the distance decreases, $f_c^{(C)} \to 1$ when $h_{i,c} \to 0$. To compute $f_c^{(C)}$, we first find $h_{i,c}$ for each company and identify the maximum distance $h_{\max}$ among all companies. We normalize by the maximum to obtain a distance that lies in the [0, 1] range with $f_c^{(C)} = 1 - h_{i,c}/h_{\max}$, so that a distance of 0 corresponds to a value of $f_c^{(C)} = 1$. We report the algorithm in the appendix. We implement the algorithm and run the experiments using Python and the NumPy, Pandas, NetworkX, Matplotlib, and Seaborn libraries. Our code and data are available at https://github.com/technometrics-lab/5-TechRank.

RESULTS

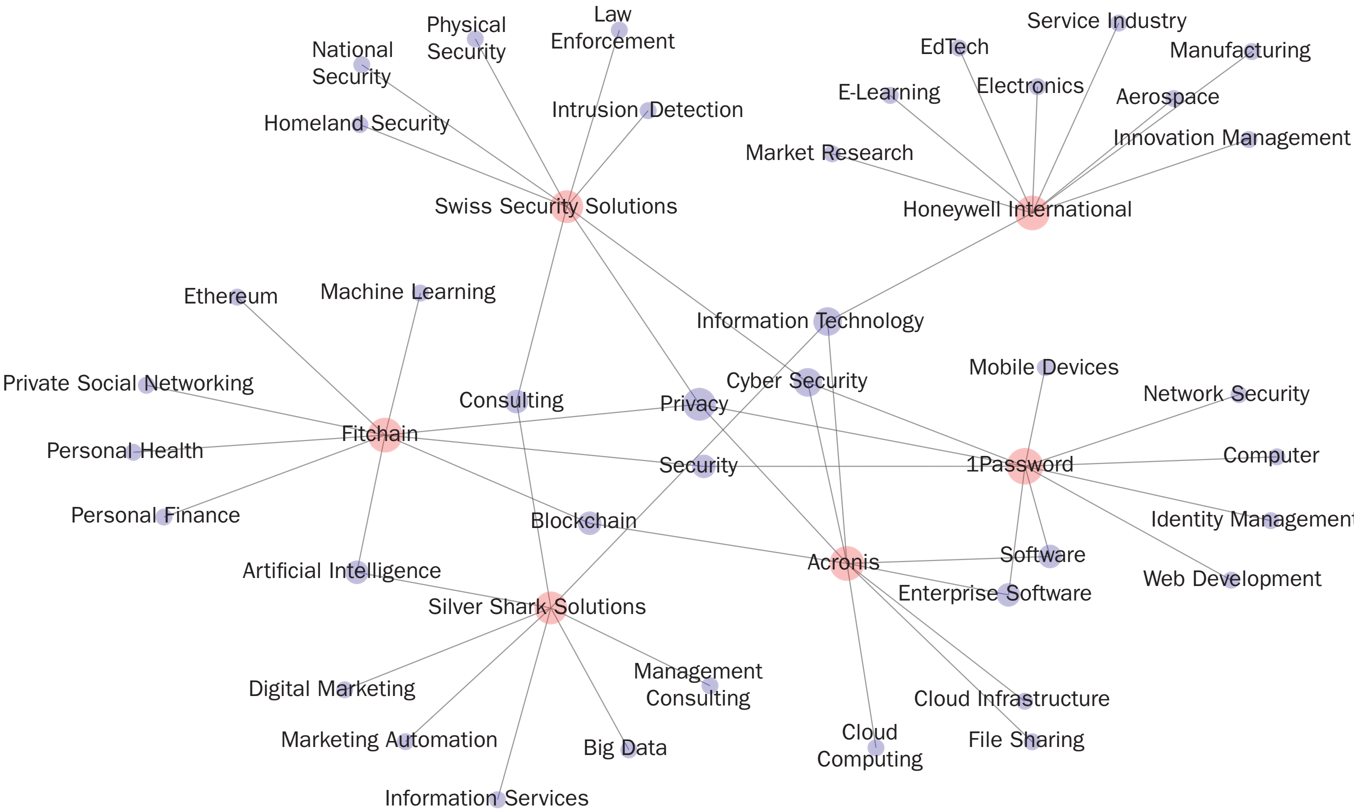
Cybersecurity Field

We select all the companies whose description contains at least two cybersecurity-related terms and obtain 2,429 companies and 477 technologies.[9] Exhibit 8 displays the structure of the bipartite network between technologies and companies.

[9]The word list is in the appendix.

EXHIBIT 8

Bipartite Network of Cybersecurity Companies



NOTES: This exhibit describes the bipartite network of a subset of cybersecurity companies (red nodes) and the technologies they are involved in (blue nodes). The node size represents the number of neighbors.

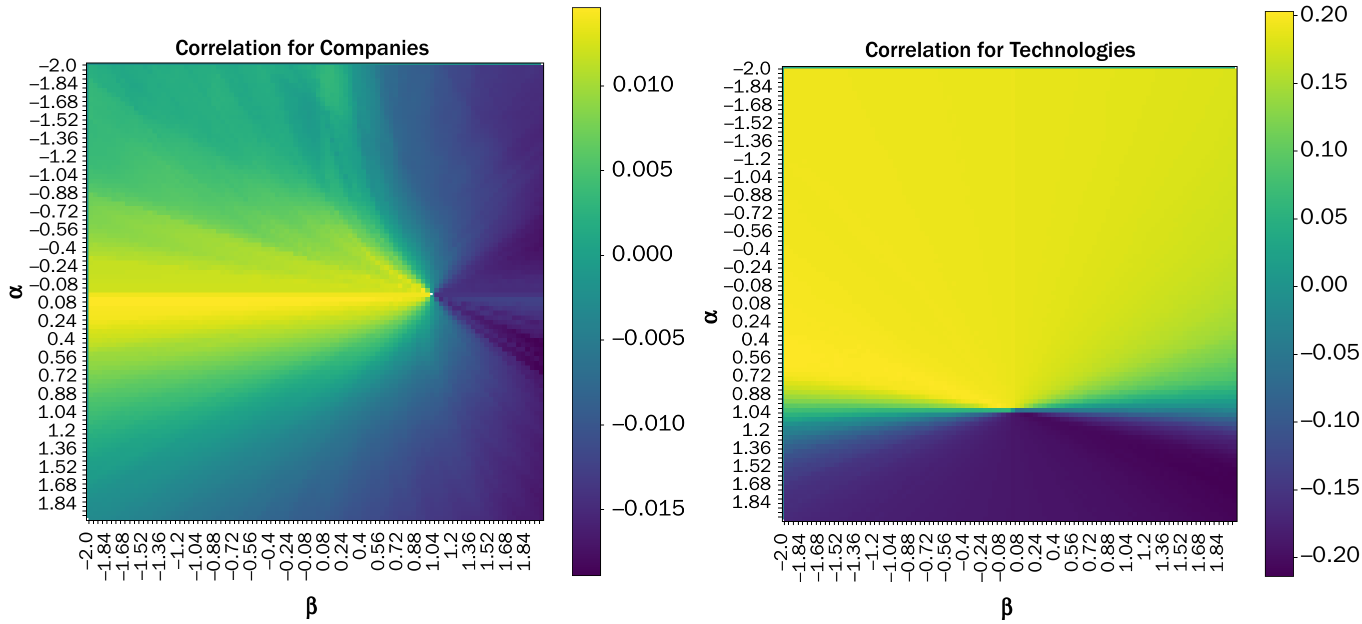
We assume investors are interested only in previous investments in technologies and companies. We examine how the parameters' calibration step changes when we change the investors' preferences using a smaller sample of companies. Exhibit 9 shows the optimization in which the correlations ρ_c and ρ_t change according to α and β. In Exhibit 10, we identify the optimal α^* and β^* to be 0.04 and –1.88 for companies and 0.48 and –2.00 for technologies, respectively. Next, we plug these values into the recursive algorithm.

We illustrate the evolution of the TechRank random walker in Exhibit 11. While the entities' positions significantly change over the first steps, they gradually stabilize. With the 2,429 companies and 477 technologies, the algorithm requires 723 (1,120) iterations for companies (technologies) to converge. Entities starting with a high score (the initialization is the degree of the node) do not significantly change rank and remain among the best ones. Thus, the algorithm assigns good scores to entities with many neighbors. Instead, entities starting with a low degree may significantly change their score, especially with technologies. TechRank not only recognizes the importance of the most established entities, but it also enables the identification of emerging technologies.

We check how TechRank performs when we change the number of companies and technologies. We fix the number of companies n^c, yielding a number of technologies n^t. For instance, in the cybersecurity field, by selecting 10 companies randomly, we

EXHIBIT 9

Grid Search of Parameters α and β



NOTE: This exhibit displays the results of the grid search for parameters α and β for 2,429 companies and 477 technologies in cybersecurity when investors' preferences are fully set in previous investments.

EXHIBIT 10

Optimal Parameters in Cybersecurity

Companies			Technologies		
Number	α*	β*	Number	α*	β*
10	–0.36	1.92	26	–2.00	0.00
100	–0.04	0.92	134	0.52	–1.04
499	–0.08	0.88	306	0.68	–1.36
997	–0.12	0.80	371	–2.00	0.00
1,494	–0.12	0.80	416	0.92	–0.12
1,990	–0.04	0.92	449	0.56	–2.00
2,429	0.04	–1.88	477	0.48	–2.00

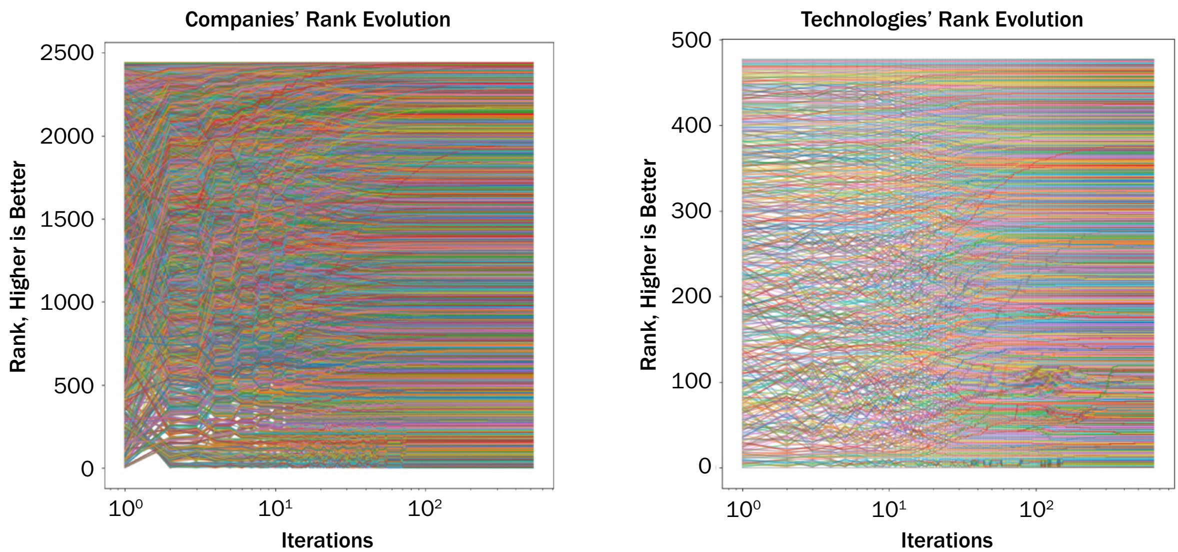
NOTE: This exhibit reports the optimal parameters α and β for companies and technologies in cybersecurity, depending on the number of companies and linked technologies considered as input.

get 26 technologies. Considering that there are 2,429 cybersecurity-related companies on Crunchbase, we study the runtime running the algorithm for 10, 100, 499, 997, 1,494, 1,990, and 2,429 companies and 26, 134, 306, 372, 431, 456, and 477 technologies, respectively.

Exhibit 12 displays the results of TechRank applied to a subset of 10 cybersecurity companies. We note that "AppOmni"'s position does not change over the iteration, while two of its technologies, "Software as a Service" (Saas) and "cloud management," increase their scores. In Exhibit 13, we display this restricted network of 10 companies, which shows that SaaS and cloud management do not have other links. Hence, the strength of this company depends on its ability to combine important technologies (software, cybersecurity, and cloud security) with more exotic fields. Similarly, "Integrity Marketing Group" is the single one involved in some fields (marketing, digital marketing, and advertising). This company does not use more-established technologies and thus does not improve its score. Again, in Exhibit 13, we observe that these technologies lie out of the main network. Conversely, "Lacework" and "Acronis" follow an opposite trend. Lacework (Acronis) significantly increases (decreases) its score. One explanation for this behavior is that Acronis is involved in many technologies, most of which are not explored by other companies. On the other hand, Lacework relies on

EXHIBIT 11

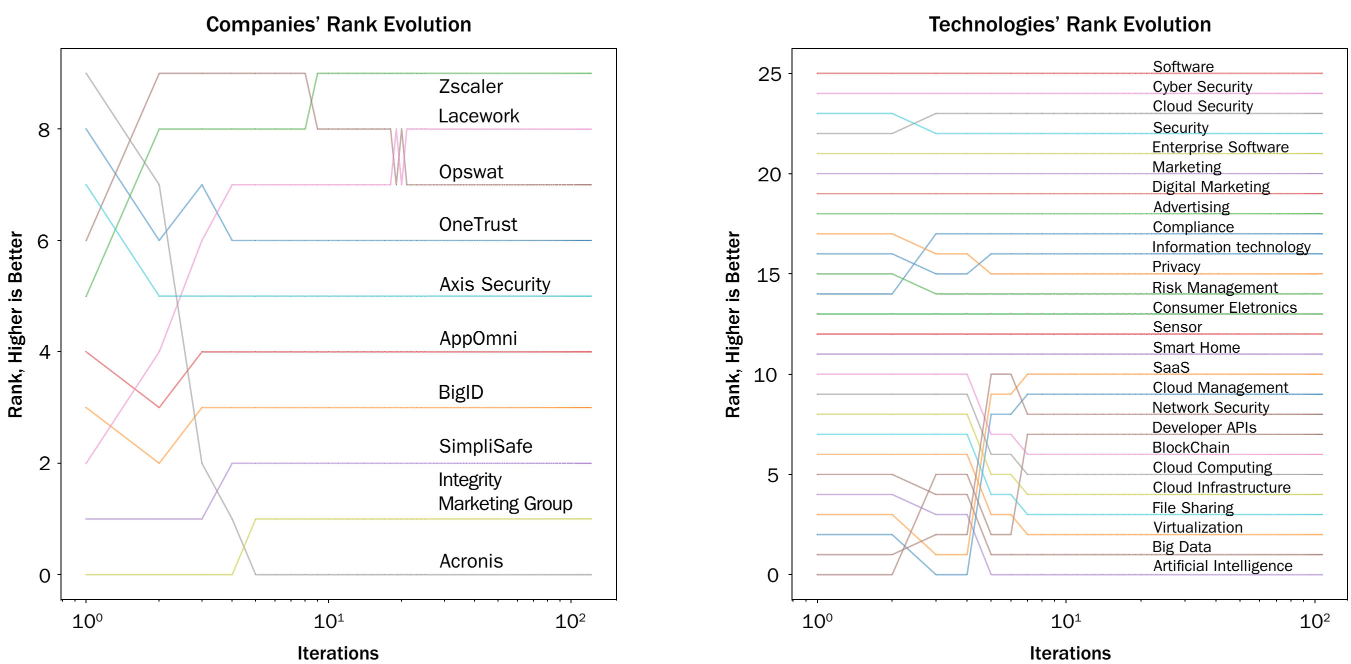
TechRank Scores' Evolution in Cybersecurity



NOTE: This exhibit displays the TechRank scores' evolution over the iterations for 2,429 companies and 477 technologies in cybersecurity.

EXHIBIT 12

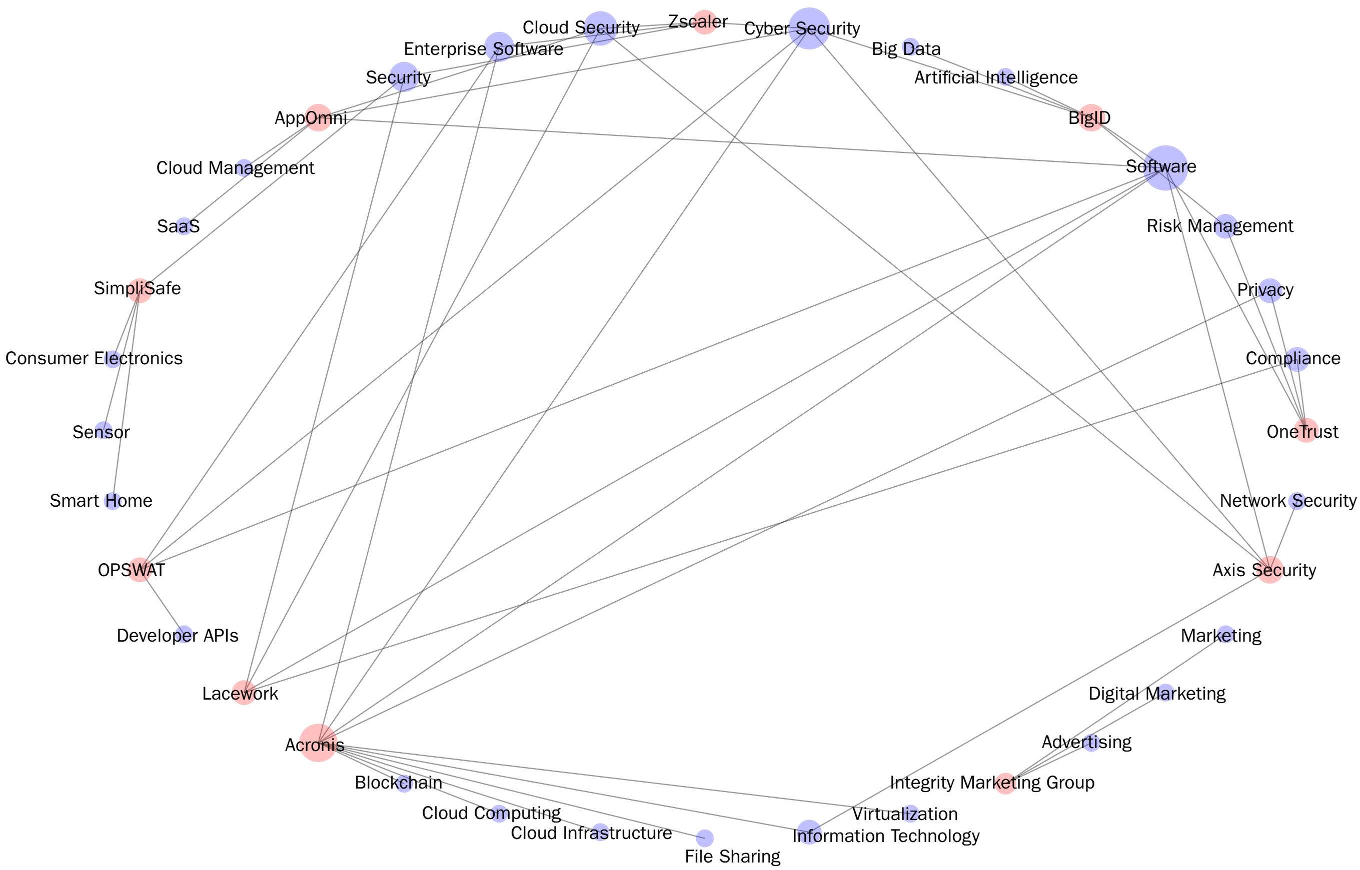
TechRank Scores' Evolution of 10 Companies in Cybersecurity



NOTE: This exhibit displays the TechRank scores' evolution over the iterations on a subset of 10 companies and 26 technologies in cybersecurity.

EXHIBIT 13

Circular Network Representation of 10 Companies in Cybersecurity



NOTE: This exhibit displays a circular network representation of a subset of 10 companies and 26 technologies in cybersecurity.

EXHIBIT 14

TechRank Convergence

Companies		Technologies	
Number	**Iterations C**	**Number**	**Iterations T**
10	32	26	18
100	100	134	155
499	134	306	2,469
997	196	371	194
1,494	180	416	871
1,990	240	449	5,000
2,429	723	477	1,120

NOTE: This exhibit reports the number of TechRank iterations before convergence for companies and technologies in cybersecurity.

recognized technologies (cybersecurity, cloud security, and software). Interestingly, compliance technology benefits from its connections, increasing its rank by three positions.

Exhibit 14 reports the number of algorithm iterations before reaching convergence. The number of iterations needed appears independent of the number of entities. Technologies need more iterations than companies, which we explain because there are many more companies than technologies. Because each company has at least one edge, the technology nodes have a higher degree than the companies, on average. Thus, we expect the structure and technology dynamics to be more complex. The algorithm complexity depends not only on the number of entities but also on the network structure.

Investment Strategy

We investigate how investors can select a strategy that reflects their preferences. If investors prefer to focus on technologies, they should choose companies working on the best technologies as selected by the (highest) TechRank score. This decision implies many criteria, such as the number of technologies they want to be invested in, the capital allocation for each company, and the diversification. We sketch the procedure to solve this decision process quantitatively in Exhibit A1 in the appendix.

Comparison with the Crunchbase Rank

Crunchbase assigns a rank to the top companies of each industry that considers the entity's strength of relationships, funding events, news articles, and acquisitions.[10] We compare our results in the cybersecurity sector with the Crunchbase rank and investigate the strength of the association between the two scores using Spearman's correlation.

We convert our algorithm's output into a ranking to make the ranks comparable. The resulting Spearman's correlation of 1.4% indicates that the two ranks are uncorrelated. We explain these differences because the Crunchbase rank is fixed, while TechRank is customizable according to investors' preferences. Moreover, the Crunchbase rank focuses on the company's activity level, not market influence. Furthermore, the Crunchbase rank results from an algorithm that involves all the companies, while we focus only on a subset. We attempt to change the investors' preferences and never obtain correlation coefficients above 2%. Other explanations for this divergence include assigning a weight identifying the distance between entities in the ranking. Along the same line, TechRank allows decision-makers to set a threshold as a starting parameter before running the algorithm. Finally, the Crunchbase algorithm is not open source, and we do not know its mechanism, which makes identifying the divergence's source difficult.

Ex Post Financial Performance

To statistically validate the ranking of our static algorithm, we regress ex post financial performance retrieved two years later on firms' rank. We use three dependent variables as financial performance metrics. First, a dummy variable takes the value of 1 if the firm has received funding within the two years following the ranking and 0 otherwise. We use a probit regression in this setting. Second, we use the total funding received by the firm in the two years following the ranking. Third, we use the latest available postmoney valuation of the firms following the ranking. We additionally use three explanatory variables—the TechRank score, TechRank firm's order in the ranking (an integer from one to the number of firms), and Crunchbase score delivered by the data provider—for benchmarking. We report our results in Exhibit 15. In the probit setting, we obtain positive and statistically significant slopes at the 5% level when the TechRank score and the firms' order are used as explanatory variables. This points to the fact that firms' ranking high in our classification subsequently have a higher probability of receiving new funding. Instead, when we use the Crunchbase rank, we obtain a negative and statistically significant slope. By the same token, when we use the continuous measure of total additional funding received subsequently to the ranking, we obtain positive but not statistically significant slopes' coefficients for both the TechRank scores and firms' order. Again, the Crunchbase score yields negative and statistically significant (at the 10% level) slope coefficients, indicating

[10] https://about.crunchbase.com/blog/influential-companies/

EXHIBIT 15

TechRank Ex Post Financial Performance for Cybersecurity Firms

	Probit Next Funding			Total Funding			PostMoney Valuations		
	TR Score	TR Order	CB Score	TR Score	TR Order	CB Score	TR Score	TR Order	CB Score
Intercept	−2.27	−2.54	−1.19	22.63×10^6	17.34×10^6	40.20×10^6	61.23×10^6	72.72×10^6	333.87×10^6
p-value/ *t*-stat	0.00	0.00	0.00	2.36	1.01	3.94	0.79	0.43	3.27
Slope	0.01	2.00×10^{-4}	−0.00	1.02×10^3	7.74×10^3	−112.33	27.56×10^6	105.21×10^3	-1.04×10^3
p-value/ *t*-stat	0.03	0.02	0.00	1.20	0.76	−1.75	4.03	1.05	−1.62
Log-Lik/ R^2	−152.76	−151.74	−117.19	0.05	0.02	0.10	0.38	0.04	0.09
# Obs.	2, 211	2, 211	2, 211	2, 211	2, 211	2, 211	2, 211	2, 211	2, 211

NOTES: This exhibit reports the results of probit (left panel) and linear regressions (middle and right panels) of (i) the occurrence of a funding round, (ii) the total funding in USD, and (iii) the latest available postmoney valuation in the two years following the TechRank static estimation. In these three settings, the explanatory variables are (i) the TechRank score, (ii) the firm's order given by TechRank (an integer between one for the worst firm and the number of firms for the best firm), and (iii) the Crunchbase score. We report the intercept and slope coefficients and their significance (p-values in the probit and *t*-statistics in the linear regression settings). We also report the model explanatory power (log-likelihood in the probit and R^2 in the linear regression settings). Over the 2,429 cybersecurity firms, we have 2,211 observations over the following two years. The study period is April 2021 for the TechRank estimation and March 2023 for the stopping date at which funding and postmoney valuations are aggregated.

that high-ranked firms on this rank obtain less funding in the future. Finally, we find a positive and statistically significant (at the 1% level) slope for the TechRank score when using postmoney valuations as the dependent variable. In this latter setting, we also reach a striking explanatory power, with a R^2 reaching 38%. We also find positive and not statistically significant slope coefficients for the firms' order. Once again, the Crunchbase score appears to be inversely related to the future firms' valuations.

Qualitative Analysis of Selected TechRank Firms

Exhibit 16 reports the aforementioned subselection of 10 cybersecurity firms ranked according to the TechRank score. This random selection is used in the beginning of this section (Exhibits 12 and 13) to visually demonstrate the functioning of the algorithm. For these firms, we add their inception year, the total funding they received, the latest available postmoney valuation, whether they are private or public, and the number of employees. The company with the top score is the single one listed, has the highest valuation, and has the largest number of employees. For the remainder of the firms, even though we do not find a monotonic decrease along all features corresponding to the TechRank score, we do identify an overall decrease in the number of employees. One outlier is Acronis, with 2,000 employees and ranking last. However, this should be put in perspective, as Acronis is also the oldest of all (20 years old) and has not made it yet to exit (IPO or acquisition). Finally, whereas the total funding can hardly be mapped to the TechRank score, the valuations we retrieve are, instead, following it closely, which is, from an investor point of view, the critical variable.

Exogenous Factors

We conduct a sensitivity analysis based on investors' preferences. We restrict the analysis to 1,000 companies, given the long runtime required. We assume investors

EXHIBIT 16

TechRank Qualitative Analysis

Firm Name	Creation Year	Total Funding	Latest Valuation	Public/Private	Number of Employees
Zscaler	2007	340	16,240 (market)	Public	4,975
Lacework	2015	1,900	8,300	Private	1,172
Opswat	2002	125	NA	Private	661
OneTrust	2016	920	5,300	Private	3,340
Axis Security	2019	99.5	500	Private (acquired by HP)	140
AppOmni	2018	123	4,300	Private	130
BigID	2016	246.1	1,250	Private	400
SimpliSafe	2006	387	1,000	Private	800
Integrity Marketing Group	2006	2,100	NA	Private	432
Acronis	2003	658	2,500	Private	2000

NOTES: This exhibit reports the subset of 10 cybersecurity firms' TechRank scores, presented in Exhibit 12 Panel A, their creation year, total funding amount (in mln USD), latest valuations (in mln USD, when available), type (public or private), and the number of employees.

EXHIBIT 17

Companies' TechRank Scores with Location

Company Rank/ Investor Location	1	2	3	4	5
New York City	New York City (USA)	Massachusetts (USA)	Quebec (Canada)	California (USA)	Singapore (Singapore)
San Francisco	California (USA)	Illinois (USA)	California (USA)	Beijing (China)	Arizona (USA)

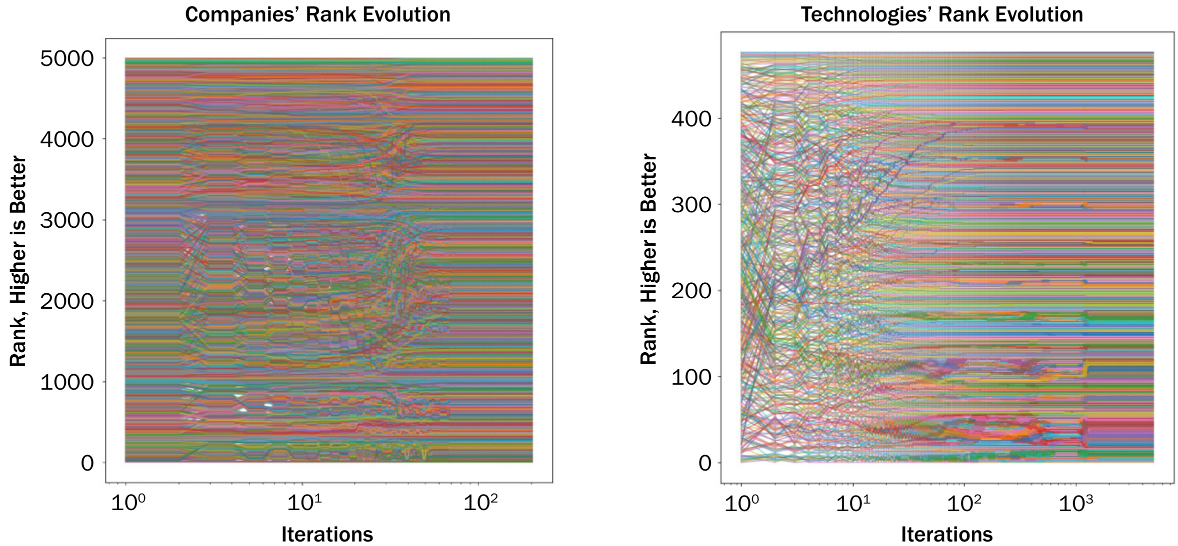
NOTE: This exhibit reports the location of the top five TechRank companies' scores when the geographical preference is fully set in the location of the investors (New York City and San Francisco).

to be interested in firm location only and consider the case of an investor based in New York City and San Francisco, in turn. In Exhibit 17, we report the outcome in terms of location for the five top-ranked companies in both cases. We uncover a location change in the company ranking, with the first being in the state of New York (investor based in New York City) and California (investor based in San Francisco), respectively. The companies with lower ranks also reflect these geographical preferences, albeit with some exceptions (Singapore and Beijing). This implies that their other attributes overcome this flaw even if remote companies are disadvantaged.

Robustness Tests: Healthcare Sector

To test the robustness of our algorithm and benchmark of the cybersecurity sector, we apply TechRank for companies in the medical sector. Given the important number of companies (twice the number of companies working in cybersecurity), we choose this sector. We select companies with the same methodology, which yields 4,996 companies and 437 technologies. Exhibit 18 shows the results of TechRank

EXHIBIT 18
TechRank Scores' Evolution in the Medical Field



NOTE: This exhibit displays the TechRank scores' evolution over the iterations for 4,996 companies and 437 technologies in the medical field.

in the medical sector. The runtime for these companies, reported in Exhibit A3 in the appendix, is on par with those of the cybersecurity sector. To make the two fields comparable, we set as an x-label the number of entities for both companies and technologies. The results reveal that the runtime of the two fields, for both the parameter calibration and the random-walker steps, follow the same behavior for both companies and technologies. Increasing the number of entities does not yield significant changes in convergence and runtime. Finally, unlike Klein et al. (2015), for which the α remains constant and β changes significantly, we observe that all of our parameters significantly change across sectors.

CONCLUSION

We introduce TechRank, an algorithm that assigns a score to companies and technologies in complex systems. This methodology constitutes the first step toward a new data-driven investment strategy, which enables investors to follow their preferences while benefiting from a quantitative approach. We include investors' preferences based on a case-by-case study. Our algorithm convergence depends on the number of entities and the complexity of the relationships within the bipartite network. Using a restricted number of companies in cybersecurity, we analyze the TechRank scores and explain entities' score variations over iterations. We use traditional ex post financial performance metrics, such as the probability of new funding rounds, total funding, and postmoney valuations. Our algorithm's score and ranking outperform the existing one readily made available by Crunchbase. We also provide additional evidence of the good performance of our ranking through a qualitative analysis of a subsample of 10 randomly selected firms. Next, we explore how results change depending on the company's location. Finally, we conduct robustness tests in the medical field, for which our results are qualitatively similar.

We believe our approach offers value by helping investors form their decisions when cash flows are not observable, thereby preventing the use of traditional discounting models. Moreover, our algorithm's flexibility allows us to include exogenous factors and preferences, which is impossible in alternative existing company ranks, such as that of Crunchbase.

Our study could be expanded in several directions. First, TechRank could be applied to sectors other than cybersecurity and healthcare. Second, it could be expanded by including additional exogenous factors. Third, given the static nature of our algorithm, it would be interesting to estimate it at different points in time and run our ex post financial performance benchmarking sequentially. Unfortunately, historical Crunchbase data are currently not available. Finally, given our algorithm performance for cybersecurity, a highly complex market, as a case study, we believe our algorithm would perform well in all markets. TechRank is a complementary, if not alternative, way to look at personalized and data-driven portfolio choices.

APPENDIX

INVESTMENT PROCESS

List of words used to identify companies' sectors: cybersecurity, confidentiality, integrity, availability, secure, security, safe, reliability, dependability, confidential, confidentiality, integrity, availability, defense, defensive, privacy.

List of words related to the medical field: cure, medicine, surgery, doctors, nurses, hospital, medication, prescription, pill, health, cancer, antibiotic, HIV, cancers, disease, resonance, rays, CAT, blood, blood transfusion, accident, injuries, emergency, poison, transplant, biotechnology, health care, healthcare, health-tech, genetics, DNA, RNA, lab, heart, lung, lungs, kidneys, brain, gynecologist, cholesterol, diabetes, stroke, infections, infection, ECG, sonogram.

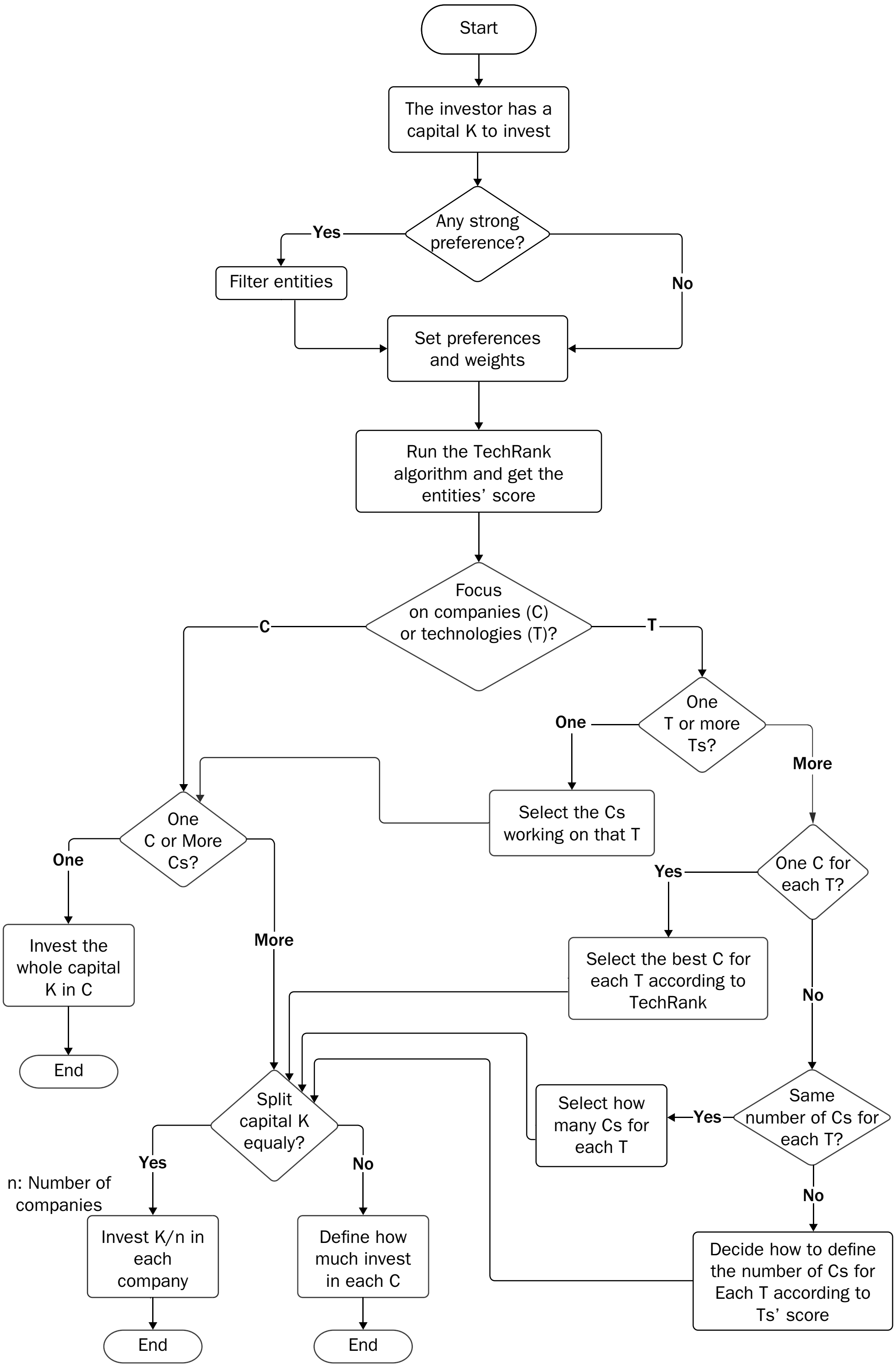
Runtime

This section provides details of the code related to the TechRank algorithm. We run it on a 16-core Intel Xeon CPU E5-2620 v4 @ 2.10 GHz with 126 GB of memory. We investigate the variations in runtime given changes in the number of companies and technologies.

The runtime is a positive function of the number of entities. For technologies, the curve is much steeper than that for companies. However, considering that the number of technologies is directly linked to the number of companies, we repeat the experiment by treating companies and technologies together. The random walk phase lines represent the runtime to convergence. There is a substantial similarity between the runtime for companies and technologies, which is surprising given their different numbers. This also shows how strongly they are correlated and supports the capability of TechRank to capture the complexity of the cybersecurity technological landscape. Exhibit A2 reports all the runtimes, and we report the corresponding runtime comparisons for technologies and companies of both the cybersecurity and medical fields in Exhibit A3.

EXHIBIT A1

Flowchart of the Investment Process



NOTE: This flowchart sketches a potential investment process that uses TechRank and investment preferences (exogenous factors and investment styles) before reaching an optimal investment and portfolio choice.

EXHIBIT A2

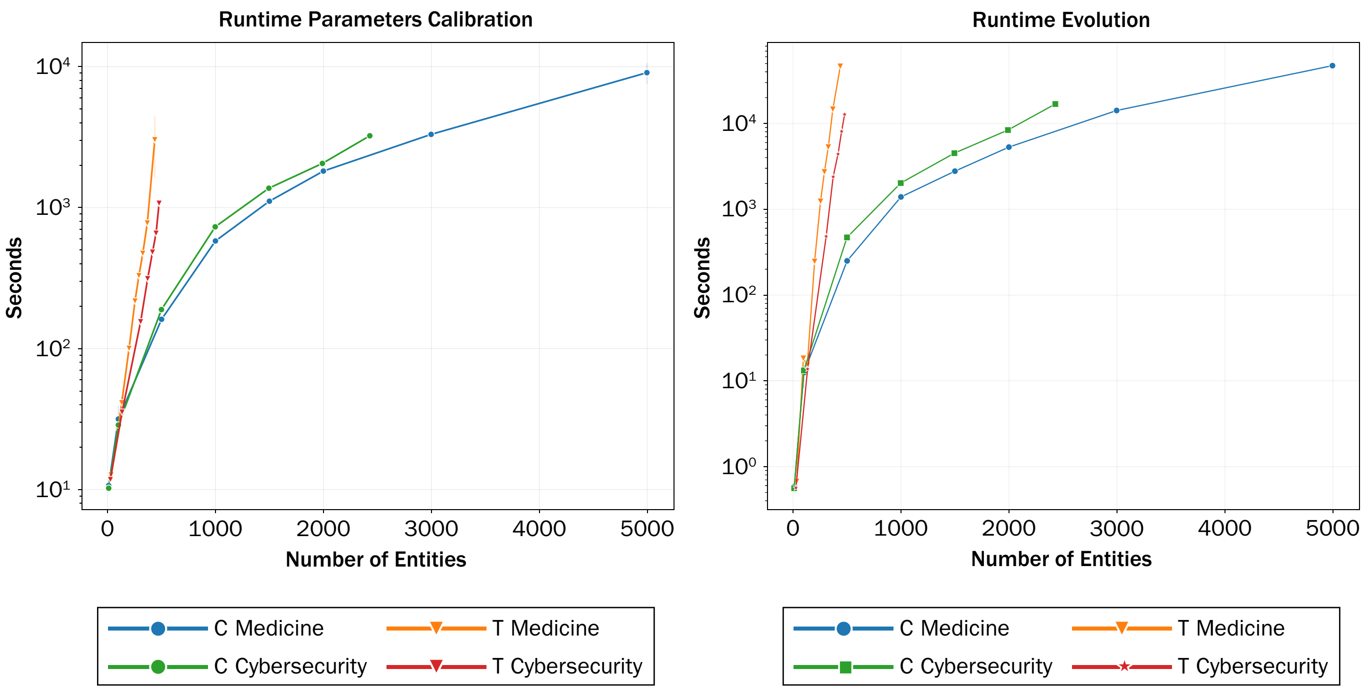
TechRank Runtime

Companies			Technologies		
Number	Parameters' Calibration	Convergence	Number	Parameters' Calibration	Convergence
10	10.21	0.56	26	11.75	0.57
100	28.69	13.24	134	35.37	13.72
499	189.03	470.10	306	154.79	483.25
997	730.43	2,023.18	371	312.65	2,392.46
1,494	1,372.17	4,514.11	416	482.18	4,404.48
1,990	2,057.42	8,396.26	449	656.95	8,096.69
2,429	3,230.99	16,890.26	477	1,071.84	12,779.62

NOTE: This exhibit reports the TechRank runtime for companies and technologies in cybersecurity.

EXHIBIT A3

Runtime Comparisons



NOTES: The left panel displays the grid search runtime for cybersecurity and medical fields. The right panel displays the parameters' calibration runtime for the cybersecurity and medical fields. The ordinate axis uses a logarithmic scale.

Generalized Equations

$$\hat{w}_c = \sum_{i=1}^{n^{(C)}} p_i^{(C)} f_i^{(C)} = p^{(C)} \cdot f^{(C)} \tag{5}$$

Considering $\sum_{i=0}^{n^{(C)}} p_i^{(C)} = 1$ and that $f_i^{(C)} \in [0,1]$ for each company i, we have $\hat{w}_c \in [0,1]$. The same holds for $\hat{w}_t$. Finally, we have

$$\begin{cases} \hat{w}_c = p^{(C)} \cdot f^{(C)} \\ \hat{w}_t = p^{(T)} \cdot f^{(T)} \\ \sum_{i=0}^{n^{(C)}} p_i^{(C)} = 1 \\ \sum_{i=0}^{n^{(T)}} p_i^{(T)} = 1 \end{cases} \tag{6}$$

where $n^{(C)}$ $(n^{(T)})$ is the number of the company- (technology-)related features and $f^{(C)} = (f_1^{(C)}, \ldots, f_{n^{(C)}}^{(C)})$.

The features related to the investments for each company are modeled as follows:

$$\begin{cases} f_1^C = \dfrac{e_{11} + e_{21}}{e_{\max}} \\ f_2^C = \dfrac{e_{22}}{e_{\max}} \\ f_3^C = \dfrac{e_{23}}{e_{\max}} \end{cases} \tag{7}$$

We define n^i (n^c) as the total number of investors (companies). Generalizing, we get,

$$\begin{cases} e_{i,c}^{IC} = \sum_{t=0}^{\mathcal{T}} \gamma_t^{i,c} & \forall i,c \\ e_c^C = \sum_{i=1}^{n^i} e_{i,c} M_{i,c}^{IC} & \forall c \\ e_{\max} = \max_c e_c^C \\ f_c^{(C)} = e_c^C / e_{\max} \end{cases} \tag{8}$$

for each c ∈ 1, …, n^c. We present the corresponding algorithm below. With Equation 8, for each company, we have a factor between 0 and 1 that summarizes the amount of previous investment.

The previous investments' factor for technology is given by,

$$\begin{cases} e_{i,c}^{(I,C)} = \sum_{t=0}^{(T)} \gamma_{i,c}^t & \forall i,c \\ e_c^C = \sum_{i=1}^{n^i} e_{i,c} & \forall c \\ e_t^T = \sum_{c=1}^{n^c} e_c M_{c,t}^{CT} \\ e_{\max} = \max_t e_t^T \\ f_t^{(T)} = e_t^T / e_{\max} \end{cases} \tag{9}$$

We provide the algorithm of this methodology below.

ALGORITHMS

Algorithm 1 Previous Investments Factor for Companies

$e^C \leftarrow [0] \cdot len(c\_names)$
for $c \in range(c\_names)$ **do**
  for $i \in range(i\_names)$ **do**
    for $c \in range(i\_names)$ **do**
      $e^{IC}_{i,c} \leftarrow \sum_{t=0}^{T} \gamma_t^{i,c}$ ▷ $\gamma_{t,c}^{t}$ is the amount of the investment from i to c at time t
      $e^C[c] \leftarrow e^C[c] + e^{IC}_{i,c}$
    end for
  end for
end for
$e^C_{max} \leftarrow max(e^C)$
$f^C \leftarrow e^C/e_{max}$ ▷ f^C: list of previous investments for each technology
return f^C

Algorithm 2 Previous Investments Factor for Technologies

$e^C \leftarrow [0] \cdot len(c\_names)$
for $c \in range(c\_names)$ **do**
  for $i \in range(i\_names)$ **do**
    $e^{IC}_{i,c} \leftarrow \sum_{t=0}^{(T)} \gamma_t^{i,C}$ ▷ $\gamma_{i,c}^{t}$ is the amount of the investment from i to c at time t
    $e^C[c] \leftarrow e^C[c] + e^{IC}_{i,c}$
  end for
end for
$e^T \leftarrow e^C \cdot M^{CT}$ ▷Matrix multiplication
$e_{max} \leftarrow max\,(e^T)$
$f^T \leftarrow e^T/e_{max}$ ▷ f^T: list of previous investments for each technology
return f^T

Algorithm 3 Geographic Coordinates Factor

$h\_dict \leftarrow \{\}$
for $c\_name, c\_address \in c\_locations$ **do**
  $lat \leftarrow c\_address \cdot latitude$
  $lon \leftarrow c\_address \cdot longitude$
  $h \leftarrow haver\_dist(lat, lon, lat\_inv, lon\_in)$ ▷ haver_dist is a function we have created
  $h\_dict[c\_name] \leftarrow 1/h$
end for
$h\_max \leftarrow \max\,(h\_dict)$
for $c\_name, h \in h\_dict$ **do**
  $h\_dict\,[c\_name] \leftarrow 1 - h/h\_max$
end for
return h_dict

Distance Computation

We obtain the distance between two points on earth with the Haversine approximation (hav(θ)), using the latitude and longitude of the locations (Ingole and Nichat 2013). Letting (λ_1, ϕ_1) and (λ_2, ϕ_2) be the longitude and latitude of two points on a sphere, and θ, the central angle given by the spherical law of cosines, the Haversine distance writes,

$$h = \text{hav}(\theta) = \text{hav}(\phi_2 - \phi_1) + \cos\phi_1 \cos\phi_2 \,\text{hav}(\lambda_2 - \lambda_1) \tag{10}$$

ACKNOWLEDGMENTS

This document is the result of a research project funded by the Cyber-Defence Campus, armasuisse Science and Technology. A preprint of this research is available on arXiv (see Mezzetti et al. 2022), which is derived from Mezzetti et al. (2021), an M.Sc. thesis conducted at the Cyber-Defence Campus. Our code and data are available at https://github.com/technometrics-lab/5-TechRank.

REFERENCES

Andries, P., B. Clarysse, and S. Costa. 2021. "Technology Ventures' Engagement of External Actors in the Search for Viable Market Applications: On the Relevance of Technology Broadcasting and Systematic Validation." *Journal of Business Venturing* 36 (6): 106145.

Baron, R. A., and G. D. Markman. 2003. "Beyond Social Capital: The Role of Entrepreneurs' Social Competence in Their Financial Success." *Journal of Business Venturing* 18 (1): 41–60.

Battiston, F., V. Nicosia, and V. Latora. 2014. "Structural Measures for Multiplex Networks." *Physical Review E* 89 (3): 1–14.

Bavelas, A. 1948. "A Mathematical Model for Group Structures." *Applied Anthropology* 7 (3): 16–30.

Benzi, M., E. Estrada, and C. Klymko. 2013. "Ranking Hubs and Authorities Using Matrix Functions." *Linear Algebra and Its Applications* 438 (5): 2447–2474.

Bonacich, P. 1972. "Factoring and Weighting Approaches to Status Scores and Clique Identification." *Journal of Mathematical Sociology* 2 (1): 113–120.

Braun, R., T. Jenkinson, and I. Stoff. 2017. "How Persistent Is Private Equity Performance? Evidence from Deal-Level Data." *Journal of Financial Economics* 123 (2): 273–291.

Brown, G. W., O. R. Gredil, and S. N. Kaplan. 2019. "Do Private Equity Funds Manipulate Reported Returns?" *Journal of Financial Economics* 132 (2): 267–297.

Brown, G., R. Harris, W. Hu, T. Jenkinson, S. N. Kaplan, and D. T. Robinson. 2021. "Can Investors Time Their Exposure to Private Equity?" *Journal of Financial Economics* 139 (2): 561–577.

Brown, G. W., K. J. Crouch, A. Ghent, R. S. Harris, Y. V. Hochberg, T. Jenkinson, S. N. Kaplan, R. Maxwell, and D. T. Robinson. 2022. "Should Defined-Contribution Plans Include Private Equity Investments?" *Financial Analysts Journal* 78 (4): 5–17.

Buldyrev, S. V., R. Parshani, G. Paul, H. E. Stanley, and H. Shlomo. 2010. "Catastrophic Cascade of Failures in Interdependent Networks." *Nature* 464 (7291): 1025–1028.

Canito, J., P. Ramos, S. Moro, and P. Rita. 2018. "Unfolding the Relations between Companies and Technologies under the Big Data Umbrella." *Computers in Industry* 99 (1): 1–8.

Christensen, C. M. *The Innovators Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business School Press: Boston, MA, 1997.

Cochrane, J. H. 2005. "The Risk and Return of Venture Capital." *Journal of Financial Economics* 75 (1): 3–52.

Crunchbase, Inc. 2022. Historical company data. Crunchbase daily *.csv export, data retrieved April 2022. https://data.crunchbase.com/docs/daily-csv-export.

Cumming, D. 2006. "Corporate Venture Capital Contracts." *The Journal of Alternative Investments* 9 (3): 40–53.

Cumming, D., L. H. Haß, and D. Schweizer. 2013. "Private Equity Benchmarks and Portfolio Optimization." *Journal of Banking and Finance* 37 (9): 3515–3528.

Dalle, J.-M., M. L. den Besten, and C. Menon. 2017. "Using Crunchbase for Economic and Managerial Research." Available at https://doi.org/10.1787/18151965.

den Besten, M. L. 2021. "Crunchbase Research: Monitoring Entrepreneurship Research in the Age of Big Data." Available at http://dx.doi.org/10.2139/ssrn.3724395.

Donato, D., L. Laura, S. Leonardi, and S. Millozzi. 2004. "Large Scale Properties of the Webgraph." *European Physical Journal B* 38 (2): 239–243.

Ewens, M. 2009. "A New Model of Venture Capital Risk and Return." http://dx.doi.org/10.2139/ssrn.1356322.

Fisch, C., and J. H. Block. 2021. "How Does Entrepreneurial Failure Change an Entrepreneur's Digital Identity? Evidence from Twitter Data." *Journal of Business Venturing* 36 (1): 106015.

Freeman, L. C. 1978. "Centrality in Social Networks Conceptual Clarification." *Social Networks* 1 (3): 215–239.

Gold, A. H., A. Malhotra, and A. H. Segars. 2001. "Knowledge Management: An Organizational Capabilities Perspective." *Journal of Management Information Systems* 18 (1): 185–214.

Gordon, L. A., M. P. Loeb, W. Lucyshyn, and L. Zhou. 2018. "Empirical Evidence on the Determinants of Cybersecurity Investments in Private Sector Firms." *Journal of Information Security* 9 (2): 133–153.

Gornall, W., and I. A. Strebulaev. 2020. "Squaring Venture Capital Valuations with Reality." *Journal of Financial Economics* 135 (1): 120–143.

Harris, R. S., T. Jenkinson, and S. N. Kaplan. 2014. "Private Equity Performance: What Do We Know?" *The Journal of Finance* 69 (5): 1851–1882.

——. 2016. "How Do Private Equity Investments Perform Compared to Public Equity?" *The Journal of Investment Management* 14 (3): 14–37.

Harris, R. S., T. Jenkinson, S. N. Kaplan, and R. Stucke. 2018. "Financial Intermediation in Private Equity: How Well Do Funds of Funds Perform?" *Journal of Financial Economics* 129 (2): 287–305.

——. 2023. "Has Persistence Persisted in Private Equity? Evidence from Buyout and Venture Capital Funds." *Journal of Corporate Finance,* forthcoming.

Hidalgo, C. A., R. Hausmann, and P. S. Dasgupta. 2009. "The Building Blocks of Economic Complexity." *Proceedings of the National Academy of Sciences* 106 (26): 10570–10575.

Hoang, H., and B. Antoncic. 2003. "Network-Based Research in Entrepreneurship: A Critical Review." *Journal of Business Venturing* 18 (2): 165–187.

Hwang, M., J. M. Quigley, and S. E. Woodward. 2005. "An Index for Venture Capital, 1987–2003." *Contributions to Economic Analysis and Policy* 4 (1): 1–43.

Ingole, P. V., and M. K. Nichat. 2013. "Landmark Based Shortest Path Detection by Using Dijkestra Algorithm and Haversine Formula." *International Journal of Engineering Research and Applications (IJERA)* 3 (3): 162–165.

Kaplan, S. N., and A. Schoar. 2005. "Private Equity Performance: Returns, Persistence, and Capital Flows." *The Journal of Finance* 60 (4): 1791–1823.

Katz, L. 1953. "A New Status Index Derived from Sociometric Analysis." *Psychometrika* 18 (1): 39–43.

Klein, M., T. Maillart, and J. Chuang. 2015. "The Virtuous Circle of Wikipedia: Recursive Measures of Collaboration Structures." In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*: 1106–1115.

Konstantinov, G. S. 2022. "Hedge Fund Networks." *The Journal of Alternative Investments* 25 (2): 14–32.

Korteweg, A., and S. Nagel. 2016. "Risk-Adjusting the Returns to Venture Capital." *The Journal of Finance* 71 (3): 1437–1470.

Kurant, M., and P. Thiran. 2006. "Layered Complex Networks." *Physical Review Letters* 96 (13): 1–4.

Liu, S. 2021. "The Role of Contracts in Venture Capital Returns." *The Journal of Alternative Investments* 24 (2): 68–88.

Mezzetti, A., D. Percia David, T. Maillart, M. Tsesmelis, and A. Mermoud. 2021. "Techrank: A Network-Centrality Approach for Informed Cybersecurity Investments." https://arxiv.org/abs/2112.05548.

Mezzetti, A., L. Maréchal, D. Percia David, W. Lacube, S. Gillard, M. Tsesmelis, T. Maillart, and A. Mermoud. 2022. "Techrank." https://arxiv.org/abs/2210.07824.

Moon, M., and E. S. Schwartz. 2000. "Rational Pricing of Internet Companies." *Financial Analysts Journal* 56 (3): 62–75.

Moskowitz, T. J., and A. Vissing-Jørgensen. 2002. "The Returns To Entrepreneurial Investment: A Private Equity Premium Puzzle?" *American Economic Review* 92 (4): 745–778.

Page, L., S. Brin, R. Motwani, and T. Winograd. 1999. "The PageRank Citation Ranking: Bringing Order to the Web." https://citeseerx.ist.psu.edu/doc/10.1.1.31.1768.

Peng, L. 2001. "Building a Venture Capital Index." http://dx.doi.org/10.2139/ssrn.281804.

Robinson, D. T., and B. A. Sensoy. 2016. "Cyclicality, Performance Measurement, and Cash Flow Liquidity in Private Equity." *Journal of Financial Economics* 122 (3): 521–543.

Saxena, A., and S. R. S. Iyengar. 2017. "Global Rank Estimation." https://arxiv.org/abs/1710.11341.

——. 2020. "Centrality Measures in Complex Networks: A Survey." https://doi.org/10.48550/arXiv.2011.07190.

Schmidt, D. M. 2006. "Private Equity versus Stocks." *The Journal of Alternative Investments* 9 (1): 28–47.

Smith, C., J. B. Smith, and E. Shaw. 2017. "Embracing Digital Networks: Entrepreneurs' Social Capital Online." *Journal of Business Venturing* 32 (1): 18–34.

Tu, X., G.-P. Jiang, Y. Song, and X. Zhang. 2018. "Novel Multiplex PageRank in Multilayer Networks." *IEEE Access* 6 (1): 12530–12538.

Xing, W., and A. Ghorbani. 2004. "Weighted PageRank Algorithm." In *Second Annual Conference on Communication Networks and Services Research*: 305–314.

Zhong, H., L. Chuanren, J. Zhong, and H. Xiong. 2018. "Which Startup to Invest In: A Personalized Portfolio Strategy." *Annals of Operations Research* 263 (1): 339–360.