TechRank: A Network-Centrality Approach for Informed Cybersecurity-Investment

Abstract. The cybersecurity technological landscape is a complex ecosystem in which entities – such as companies and technologies – influence each other in a non-trivial manner. Understanding influence measures of each entity is central when it comes to take informed technological investment decisions for ensuring critical-infrastructure protection. To recognize the mutual influence of companies and technologies in cybersecurity, we consider a bi-partite graph that links companies and technologies. Then, we weight nodes by applying a recursive algorithm based on the method of reflection. This endeavor helps to assign a measure of how an entity is impacting the cybersecurity market. Our results help (i) to measure the magnitude of influence of each entity, (ii) decision-makers to address more informed investment strategies, according to their preferences. Finally, a research agenda is suggested, notably for making a tailor-made investment by weighting specific features of entities.

Keywords: Critical infrastructure · technology monitoring · bipartite networks · centrality measure · optimal investment.

1 Introduction

The operational continuity of critical infrastructures (CI) is central for the functioning of modern societies. Yet, CIs are managed and monitored by interdependent information systems, exposing CIs to cascading failures [5]. In this extreme risk context [1], the security of information systems is of crucial importance [9]. However, developing and implementing an effective defense for information-systems is a non-trivial task. Indeed, the cybersecurity market is shaped by complex and fast-paced technological developments. Research and development is a costly and risky activity that may lead to sunk costs and poor development prospects. Hence, uncertainties related to cybersecurity technologies – and companies that develop them – are ubiquitous [1]. As a matter of fact, around 90% of startups fail. In 42% of cases, such failures are due to market-demand misreading, while in 29% of cases, startups fail because they run out of funding [11]. The decision-making process associated to cybersecurity investment and procurement is no exception, and is thus a challenge.

Our objective is to contribute in helping CI decision-makers to make more informed-decisions for investments in cybersecurity. For that purpose, we first

[1] The hack of the Colonial Pipeline in May, 2021 is one of the most significant cyberattack on a national CI in history. This case shows the extreme risk context and hence the importance of making informed technological investments for the (cyber) security of CI. Source: https://www.bbc.com/news/technology-57063636

model and map the ecosystem of entities (i.e., technologies, companies) from *Crunchbase*. This mapping reflects the cybersecurity market through a bi-partite network. Then, we evaluate their relative influence in the whole ecosystem by adapting a recursive algorithm that returns a network-centrality measure. This should help decision-makers and investors to quantitatively assess the influence of the entities that constitute the cybersecurity ecosystem, thus reducing potential investment uncertainties and optimize the procurement process.

The remainder of this paper is structured as follows: Section 2 present the related work; Section 3 presents the data and methods; Section 4 shows the preliminary results; Section 5 sets the agenda for future works and discusses limitations; while Section 6 concludes.

2 Related Work

Network-centrality measures – i.e., measures that assess the importance/influence of nodes in networks – have been widely investigated (see [2,7] for extensive literature reviews in the field of social networks and complex systems). Some extensions and re-adaptations of the well-know Google's *PageRank* [8] algorithm have been investigated, including for bi-partite networks. For instance, Hidalgo et al. [4] developed the method of reflection: an algorithm that characterizes the structure of bi-partite graphs by iteratively calculating the mean value of previous-step properties of a node's neighbors. Similarly, previous work by Klein at al. [6] assesses the relationship between editors and articles on Wikipedia by extending the *PageRank* centrality measure. They develop a recursive algorithm to measure how editor expertise influences the quality of articles, and how contributions to articles influence editors' expertise. Their work shed some light on a cooperation and coordination scheme, which aims to solve common interaction problems that emerge within such structures.

To the best of our knowledge, despite a prolific literature related to economic valuation for cybersecurity investment (see [10] for an extensive literature review), little work has been done when it comes to assess the importance/influence of entities in a network, especially for the field of cybersecurity and CIs. The idea of measuring the global rank of a node starting from local information is still an open research gap. Hence, in this work, we use a similar approach to the one of Klein et al. to investigate the following research question: how to measure the influence of an entity, according to its relations in the graph, in order to quantitatively design the best investment strategy?

3 Data and Methodology

We use a Crunchbase[2] (CB) dataset to build our bi-partite network model, which is composed of two node types : companies and technologies. CB encompasses information about companies activities and their financing by leveraging big

[2] https://www.crunchbase.com/; data downloaded on April 28th, 2021.

data and open-source information in a semi-automated fashion. Data are sourced thanks to investors and the community of contributors. As emphasized by Dalle et al. [3], CB is widely used by researchers for the quality of its data and the usability of their structure. For the sake of brevity, we refrain from a detailed dataset description which can be found on the CB enterprise API. [3]

Figure 1 gives an idea of the structure of the bi-partite network that describes on which technologies each company is working on.

We adapt the recursive algorithm developed by Klein at al. [6], based on the method proposed by Hidalgo et al. [4]. This method should encompass the complex structure of cooperation and competition of the technological landscape. The resulting rank condenses the positive influence of experienced companies on technologies and, together, the positive impact of newborn companies on novel fields. In the same way, a company is going to receive a higher rank thanks to the positive influence of important technologies. Klein et al. [6] claim that too many editors working on an article can create disvalue. We investigate if the same applies for cybersecurity technologies: if too many companies work on the same field, the business gap will narrow, and companies will lose market share.

Starting from Klein at al. [6], we build an adjacency matrix $M_{c,t} \in R^{N_c,N_t}$ that takes value 1 if a company c works on a technology t and 0 otherwise. N_c represents the total number of cybersecurity companies we are considering and N_t the total number of technologies. The aim of the algorithm is to assign a a weight to every node, which sums up its relevance within the graph: a contribution value to each company and a quality value to each technology. The starting

[3] https://app.swaggerhub.com/apis-docs/Crunchbase/crunchbase-enterprise_api/

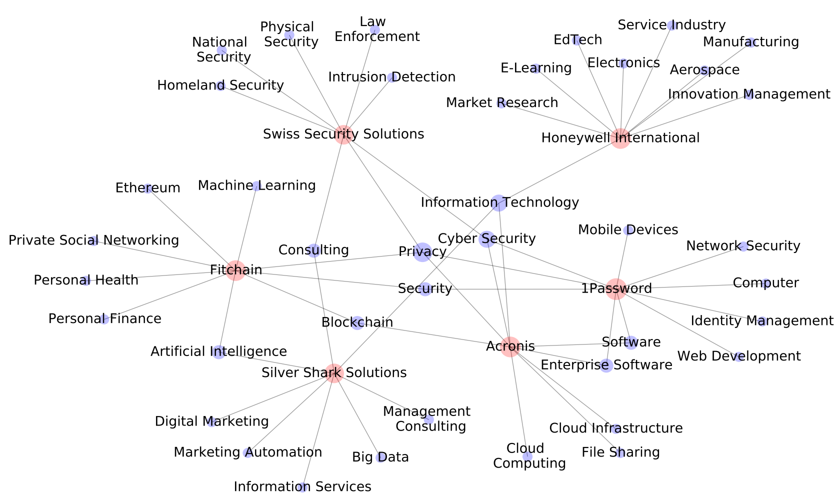

Fig. 1: Bi-partite network for selected cybersecurity companies (red nodes) and technologies (blue nodes) they are working on. The nodes' size represents the number of neighbors.

point consists in measuring the expertise of a company (w_c^0) by summing the number of technologies it works on: $w_c^0 = \sum_{t=1}^{N_t} M_{c,t} = k_c$. The same holds for technologies: $w_t^0 = \sum_{c=1}^{N_c} M_{c,t} = k_t$.

First, we recall that this algorithm is a Markov process: the step w^n depends only to information available at the previous one w^{n-1}. At each step, the method incorporates information about the expertise of companies and the relevance of technologies, leveraging the bi-partite network structure. The whole process can be seen as a random walker that jumps with a transition probability that is zero in case $M_{c,t} = 0$. We need to define two matrices that explain how we move from one step to another one: they represent the probability of jumping from technology t to company c and depends on the initial conditions.

$$\begin{cases} G_{c,t}(\beta) & = \frac{M_{c,t} k_c^{-\beta}}{\sum_{c'=1}^{N_c} M_{c',t} k_c'^{-\beta}} \\ G_{c,t}(\alpha) & = \frac{M_{c,t} k_t^{-\alpha}}{\sum_{t'=1}^{N_t} M_{c,t'} k_t'^{-\alpha}} \end{cases} \tag{1}$$

Klein at al. [6] also introduce two parameters, α and β, that inform how coordination generates value. Thanks to $G_{e,a}$, we get the recursive step:

$$\begin{cases} w_c^{n+1} & = \sum_{t=1}^{N_t} G_{c,t}(\beta) w_t^n \\ w_t^{n+1} & = \sum_{c=1}^{N_c} G_{c,t}(\alpha) w_c^n \end{cases} \tag{2}$$

Similarly to *PageRank*, the recursion ends when the rank stabilizes.

The CB platform assigns a rank to the top companies – according to their algorithm – in each industry. The CB rank takes into account the entity's strength of relationships, funding events, news articles, acquisitions, etc.[4]. We compare our results in cybersecurity with this rank. We investigate the strength and direction of the association between the two scores using the Spearman's rank correlation coefficients.

4 Preliminary Results

Our fist results are based on the selection of all companies whose description contains at least two words related to the field of cybersecurity. We get a total of 2,443 companies and 478 technologies. Figure 2 shows that the recursive algorithm introduced in Section 3 converges for both companies and technologies after a sufficient number of iterations.

As mentioned in Section 3, we compare our ranking with the CB rank, in order to get a baseline. Thus, to make the ranks comparable, we convert our algorithm's output into a ranking. The resulting Spearman's correlation (0.014) shows that the two ranks are not correlated: even if the goals of the *TechRank* and the CB rank are similar, this outcome reflects their substantial differences. First, we do not know the exact mechanism by which CB (which is not open

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

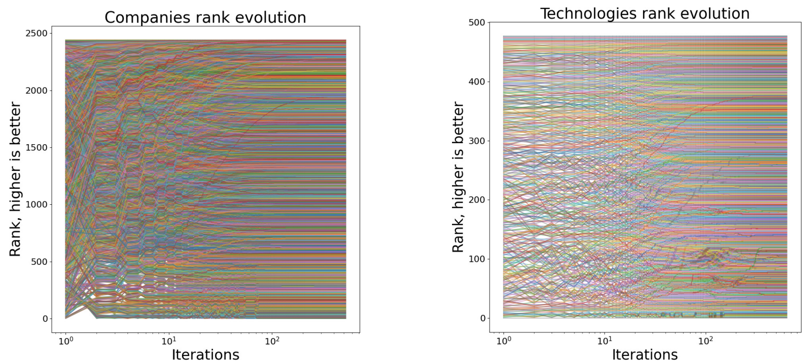

Fig. 2: Convergence evolution of the rank algorithm. The graph shows that converges in 412 iterations for companies and in 538 for technologies.

source) ranks the entities and thus potential investors cannot understand how much this rank reflects their preferences. Moreover, the CB score focuses more on the level of activity of the company, rather than its influence on the market. Among all the factors that influence it, nothing can be found about if and how technologies contribute to the results which, on the contrary, is really important in our case. Moreover, CB simply ranks companies, while our algorithm assigns a weight, which allows to identify not only the order, but a quantitative idea of the distance between one entity and the next one. These discrepancies lead to different results and we believe that our less-is-more approach, which directly depends on the capabilities of the companies, is a good way for investors to analyze if an entity has good development prospects. This, together with the next steps explained in Section 5, leads to a personalizable and see-through approach, new compared to the common un-adaptable rankings currently available.

In particular, analyzing the singular entities, we see that new fields have a good impact on companies working on them and that too much competition on a technology has a negative impact on its neighbors. Another benefit is the opportunity of ranking also technologies: some investors may be interested in a certain field and not in companies only. The *TechRank* methodology enables them to create the portfolio that better reflects their preferences.

5 Further Steps

Our research agenda will focus on two steps: first, including the influence of exogenous factors –such as impact of incubators and social aspects– and, second, creating the optimal portfolio strategy. The first step incorporates also the impact of previous investments, captured by another bi-partite structure composed of investors and companies: each company is linked to its investors and edges are weighted by the investment amount. Tracking the investments is a relevant factor to "follow the money" and describe how much a given entity has been trusted by investors. Once we have defined the entities to invest in, for the last step, we use modern portfolio theory, which essentially consists in maximizing returns while minimizing variance.

The outcome should allow investors to personalize the algorithm according to their preferences. As a matter of fact, the main two strengths of *TechRank*, compared to currently available ranks -as the CB one-, are its customizable and transparency features.

6 Conclusion

The aforementioned methodology constitutes the first step towards a new data-driven investment strategy, which allows investors to follow their preferences while benefiting from a quantitative and data-driven approach.

Thanks to the interdisciplinary core of this solution, investors can undertake a transparent decision making process when dealing with highly complex scenarios, as the cybersecurity market. In finance, the efficacy of both "classical" technical and fundamental analysis is disputed by the efficient-market hypothesis. Our *TechRank* is the kickoff for a complementary (or even alternative) modern technological portfolio analysis for CI operators. Thus, as we develop the methodology in the CI domain, we believe that the algorithm is extendable to every large organization dealing with high level of uncertainty. Therefore, we expect that the *TechRank* algorithm will be applied and further developed in other fields.

References

1. Anderson, R.: Why information security is hard - an economic perspective. In: Seventeenth Annual Computer Security Applications Conference (2001)
2. Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.G., Petri, G.: Networks beyond pairwise interactions: structure and dynamics. Physics Reports (2020)
3. Dalle, J.M., den Besten, M., Menon, C.: Using Crunchbase for economic and managerial research. Tech. rep. (2017)
4. Hidalgo, C.A., Hausmann, R.: The building blocks of economic complexity
5. Hines, P., Talukdar, S., et al.: Controlling cascading failures with cooperative autonomous agents (2007)
6. Klein, M., Maillart, T., Chuang, J.: The virtuous circle of wikipedia: Recursive measures of collaboration structures (2015)
7. Landherr, A., Friedl, B., Heidemann, J.: A critical review of centrality measures in social networks. Business & Information Systems Engineering
8. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Tech. rep. (1999)
9. Percia David, D.: Three Articles on the Economics of Information-Systems Defense Capability. Ph.D. thesis, Université de Lausanne (2020)
10. Schatz, D., Bashroush, R.: Economic valuation for information security investment: a systematic literature review. Information Systems Frontiers
11. Steward, J.: The ultimate list of startup statistics for 2021 https://findstack.com/startup-statistics/