POLITECNICO DI TORINO

Corso di Laurea in Physics of Complex Systems

Tesi di Laurea Magistrale

Business strategies in Darknet marketplaces

An attempt to model competition in the framework of Economic Complexity

Relatore
Prof. Luca Dall'Asta

Correlatore:
Prof. Didier Sornette

Gian Giacomo della Torre
matricola: 240721

Anno accademico 2017 – 2018

Acknowledgements

I would like to take this opportunity to thank all the people who have helped me in the completion of this Master's Thesis. First of all, I am grateful to my supervisor at the Future Resilient Systems – Singapore-ETH Centre | ETH Zurich, Prof. Dr Didier Sornette, for all the time and the patience he has devoted to me. Many thanks also to my closest guidance, Dionysios Georgiadis, who helped me throughout these past few months with constant motivation, enthusiasm and great humanity. I would also like to express my gratitude to Dr Ravi S. Kudesia and Dr Sandro Lera, who, despite the little time spent together, have always been available to answer to my doubts, providing me with valuable inputs and advice.

Summary

What are the sources of competitive advantages of vendors in an economy of purely rational and anonymous agents? Hereafter, throughout the case study of Darknet markets, we attempt to develop a new model of competitive advantage that brings together knowledge of strategic management and economic complexity theories.

Over the last five years, researches in the field of Darknet marketplaces have been started and a partial light has been shed over the mystery of the illegal e-commerce. Ultimately, as a first approximation, a Darknet market can be defined as a virtual tripartite network of anonymous vendors, products and customers. In this context, the Darknet market constitutes a good approximation of an economy of purely rational and anonymous agents, with limited contacts and relationships, based only on economic exchange. In other words, the first approximation of a "textbook economy". Given such a general opportunity, it is legitim to question if is it possible to develop an economy of agents who do not directly trust and know each other? Moreover, if it's that the case, what are the attributes that make some vendors develop a competitive advantage despite their anonymity?

In this Master's Thesis, we indagate such questions through an innovative methodology that merges together literature of strategic management and economic complexity theories. Barney's model of sustained competitive advantage states that different vendors in the same market can reach different performance levels and that these differences in performance are driven by the resources that vendors possess. More in details, we show that these ideas at the vendor level closely align with Hidalgo and Haussman's theory of capabilities, which considers the set of different products as the source of competitiveness at the national level. Hence, reshaping Barney's model into the Hidalgo and Haussman theory of capabilities, we are able to propose an analysis of our Darknet dataset, and more generally, of our "textbook economy".

Starting from these arguments, in order to tackle the analysis quantitatively, we developed a generalization of the Fitness-Complexity Algorithm of Pietronero et al. Namely, we proposed a new metric over a tripartite network of vendors, products, and customers, that considers the ability to diversify over a heterogeneous set of customers as the ultimate resource of vendors' competitiveness. Results show a clear correlation between the so-defined vendors' fitness and vendors' revenues, hence they support the informativeness of the new metric.

Concluding, we attempted to describe sources of competitive advantages of a general textbook economy, throughout the case study of Darknet markets. In order to do so, we connected literature of strategic management and economic complexity and we introduced a successful tripartite generalization of the Fitness-Complexity Algorithm which quantifies

the competitiveness of vendors.

Our contribution is thus twofold: not only we have successfully analyzed the strategic dynamics of vendors in the Darknet markets, but, starting from this stimulus we have been able to develop a broader approach that enlarges the applicability of the economic complexity theories to an individual level, and, potentially, could successfully uncover this fitter-gets-richer phenomenon in many other tripartite competitive systems.

Contents

List of Tables

List of Figures

Introduction

In the course of the past few years, research in the field of Darknet marketplaces has been started. The increasing number of users and new markets in the area of illegal e-commerce have attracted the attention of both institutions and researchers. The first attempts of exploratory research were mainly focused on extracting the principal features of the markets, such as the size of transactions and the typology of products that have been exchanged. However, is still poorly under-examined what are the principles of evolution of such markets, and what role the agents' anonymity plays in the market dynamics.

Consequently, the aim of this master thesis is to shed a light on the interactions and positive synergies that take place between the economic agents of the Darknet marketplaces, as a reflection of their anonymous activities. In particular, we analyze the dynamics of vendors, products and customers, and try to address what determines the success of some vendors among others and what are the attributes that better shape their competitiveness. Moreover, Darknet markets can be considered as a case study of a general economic network; a set of tripartite relations between purely rational and anonymous agents mediated through the exchange of products. Such a framework gives the opportunity to apply and elevate the results of the analysis to the level of a general economy. In other words, the first approximation of a "textbook economy".

We started the analysis with a review of the literature of the Darknet marketplaces, understanding what ensures the level of trustworthiness between the agents in such an anonymous and considerably high-risk environment. Making use of the literature and after a preliminary and exploratory data analysis of the available datasets, we decided to approach the investigation with tools that brings together economic complexity theories and literature of strategic management. Firstly, we have adopted the theory of intangible capabilities as our key assumption. We have based the competitiveness of the vendors on the complexity of the products that they list, namely, we connected their potential of growth just with the set of products that they are able to list in the market. Subsequentially, the alignment between Barney's model of competitive advantage and theory of intangibles capabilities permitted us to scale arguments, applied on a national level, down to an organizational/individual level. The use of such a theoretical assumption reflects, methodologically, on the possibility of coherently apply the Fitness-Complexity Algorithm on the Darknet market's vendors-products network. Furthermore, in this thesis, we have understood some of the inflexibilities of the Fitness-Complexity Algorithm due to its strict requirements in the shape of the dataset. Unfortunately, the particular market circumstances, dictated by the anonymity of the agents, their increasing number and the increasing number of products, make the application of the algorithm non-trivial and to some extent irrelevant. It seemed that in this economic context, the theory of capabilities, related only to the products listed by the vendors, failed to discount all the information on the vendors' capabilities and ultimately their competitiveness.

Thereupon, we have conjectured that part of this incoherence is due to the lack of the customers' information. In fact, the original application of the Fitness-Complexity algorithm forced us to take into account a bipartite network, in our case the vendors-products network. Due to this reason, it seemed that in this circumstances the algorithm blindly concentrated the analysis just on how the market offering shapes itself, without considering its relations with the market demand. We have, finally, retained necessary to enlarge the analysis to a tripartite network of vendors, products and customers and, thus, to propose a coherent generalization of the algorithm and the theory of capabilities itself. This ultimate step has required the introduction and the testing of new metrics never attempted in the field of economic complexity.

Summarising, the results of the economic complexity approach underline the specific circumstances required by the Fitness-Complexity algorithm convergence, disproving its flexibility. Moreover, the discrepancies in the convergence behaviour of the algorithm with respect to the case of the global economy underline the differences between a mature and the stable national market with respect to a new roaring economy of vendors, customer and products, such as an online Darknet marketplace.

Chapter 1

Darknet marketplaces

1.1 General information about the Darknet marketplaces

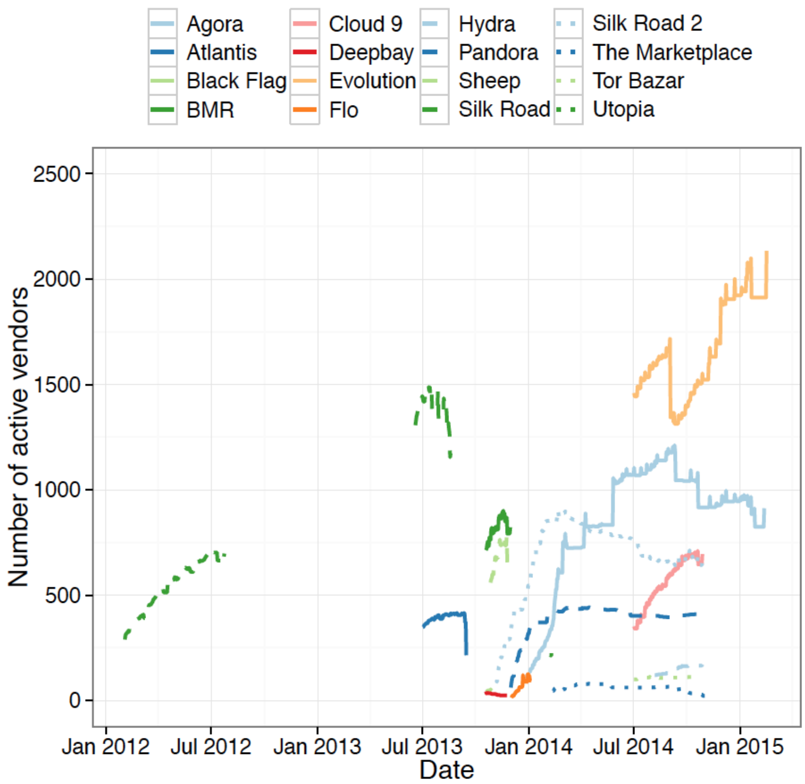
In early October 2013, the most prominent Darknet market place for the trafficking of illegal goods, known as the Silk Road, was shut down and its operator arrested. By early November 2013, a novel incarnation of the Silk Road, called "Silk Road 2.0" had been established and in a few months, numerous marketplaces appeared. All of them attached to the same standard definition of Darknet market, as an anonymizing network where connections are made only between trusted peers, using non-standard protocols and ports, accessible only by the Tor Network Hidden Services and adopting Bitcoin for any transaction. Between 2013 and 2015, the number of markets increased further, up to a number that varies from 20 to 35 or more. A recent work [1] by Dr Kyle Soska and Prof. Nicolas Christin proposed the first scraping attempt of the Darknet activity, yielding to a dataset of 3.2 TB in size and 35 different marketplaces. The paper, [1], analyses the evolution of the main illicit platform, in term of their number of transactions and number of daily volume, and, it also clears out that Silk Road, Silk Road 2, Agora, Evolution, BMR, and Hydra were the main traffic platforms, between the period 2011-2015. These remarkable results are shown in Figure 1.1. For the purposes of this master thesis, we will focus mainly on the dataset related to Silk Road and Agora. These two different data sets give us the opportunity to understand if the evolutional features are singular for each market or if instead, there are some common properties.

1.2 Trust in the Darknet marketplaces

At the beginning of our study, we believed that Silk Road's and Agora's datasets were ideal to study the formation of trust in a high-risk environment. We thought that ordering an illegal good required that the customer needed to trust the seller and, "What makes one party trust another in a high-risk situation?" seemed to be a reasonable question.

Unfortunately, the literature revealed us that at the base of any Darknet marketplace of first generation there was a centralized escrow system, used to mitigate counterparty risk [3]. In its purest form, an escrow simply means a deed held in trust by a third party until

Figure 1.1: Darknet Markets evolution in terms of number of transactions



Darknet Markets evolution in terms of number of transactions, [1].

a further condition is satisfied. The Silk Road escrow system is illustrated in Figure 1.2. As mentioned above, the most popular payment currency used on DNMs was Bitcoin (BTC), and, on the Bitcoins' blockchain, every payment was ordinarily managed through an escrow system which provides buyers and vendors with security and confidence in the validity of transactions. The escrow system and the use of the Bitcoin currency ensured that transactions were fully and safety completed to all parties satisfaction. Basically, the risk of buying and selling anonymously was mitigated by the trust on the system process, and in the administration of the market. A recent research, from the Global Drug Policy Observatory, [3], shows that the trust in the Silk Road centralized escrow system was very high no matter all the risks connected with transacting illicit goods. As an example, as one user in the data suggested ". . . the beauty of escrow is that you don't have to trust them [vendors] I often try out new vendors, they are usually eager to impress you with their wares and service". One of many other quotes of this kind [3].

In addition to the escrow system, the Darknet market gave the possibility to customers of releasing a 1 to 5 stars feedback after each transaction, in order to ensure, even more, the level of trustworthiness. It has been shown how this and the market's forum were the central mechanisms that permitted the trust to be discursively constructed in Darknet markets, [4]. Indeed, even the presence of a market forum played a key role in a safety and democratic management of the market activities. In fact, in some circumstances, the escrow system was not enough to protect the economic agents. It was, for example, the common case in which a vendor explicitly required customers to "Finalize Early" before receiving products. This practice involved customers allowing the escrow system to release the funds to the vendors, prior to the product's delivery, in good faith. Customers might later edit their review appropriately to match their actual experience but it has been observed that it was common for a buyer to give a 5 stars feedback if it has decided to finalize early, probably to minimize the likelihood of a not carried out transaction [5]. In this cases, the market's forum has permitted a severe obstruction of the non-loyal vendors, ensuring the high level of common trust between the economic agents.

Figure 1.2: Silk Road Escrow System

Silk Road Escrow System, [3].

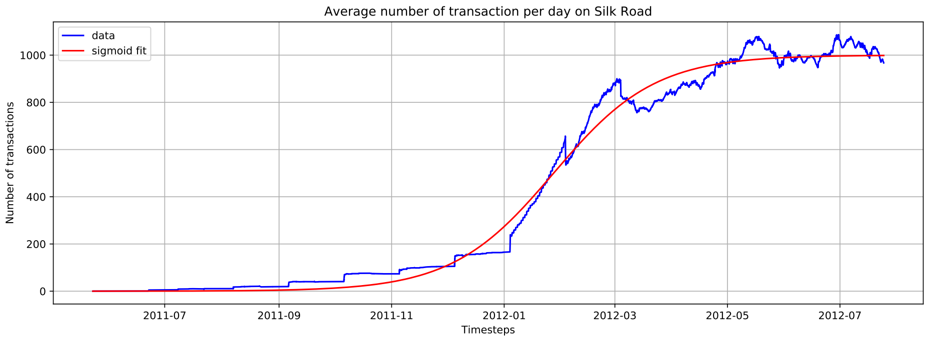
1.3 Exploratory data analysis

It is worthy to notice that to arrive to embrace a specific research thesis we have familiarized and queried the datasets in different ways, the most insightful are presented in the following paragraphs.

1.3.1 General analysis of Silk Road

In the Silk Road dataset, we have data on operations over a time window of approximately one year, from July 2011 until July 2012, agents have facilitated 190k transactions, with a total of 1100 active vendors and 160k active buyers, [2]. The daily volume has been of approximately 1k transactions a day at its peak and it enabled the trafficking of 300k USD worth of illegal goods daily. Figure 1.3 displays the daily average number of transactions, over a two-week rolling window. We can see that the adoption of Silk Road roughly follows a sigmoid function, consistent with the model of product adoption. This implies that, by the time Silk Road was dismantled, it had already reached a relatively mature stage, we suppose that it had just surpassed the first market adoption of the main products shown in Figure 1.6.

Figure 1.3: The Silk Road's adoption curve



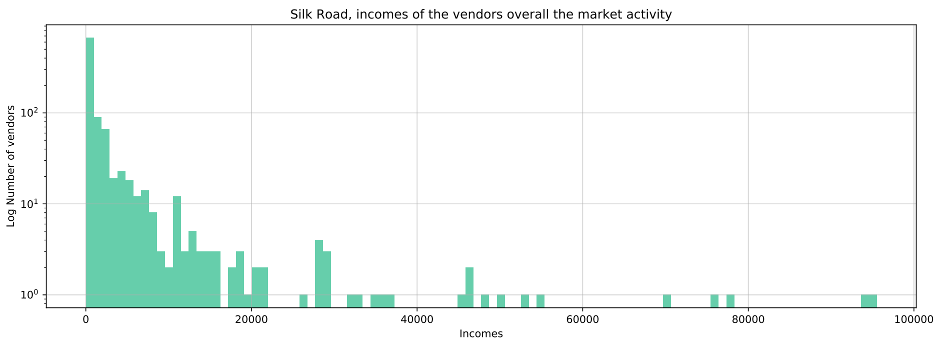
The plot has been realized taking the average number of transactions per day over the Silk Road DNM, smoothened over a two-week rolling time window. The red line is a sigmoid fit that has been implemented over the time series, it implies a typical product adoption behaviour. The staircase-like behaviour on the left side of the plot is most likely attributed to problematic data.

1.3.2 Silk Road vendors analysis

Vendors' incomes, some notable case of success

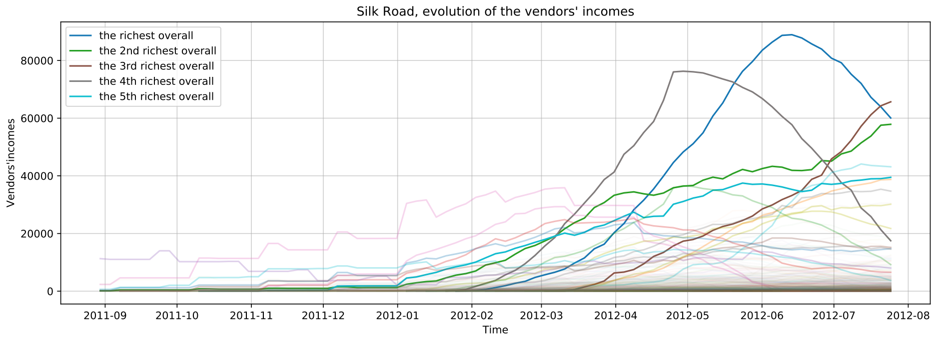
In the Silk Road dataset, we can appreciate some singular successful vendors. Figure 1.4 and Figure 1.5 show that some of them have been able to consistently dominate the market with incomes over the average. They have been able to constantly beat the competition in a market that was continuously growing in terms of products diversification and basin of customers. The Darknet market literature has underexamined what were the strategies adopted by these vendors and where their strength resided.

Figure 1.4: Silk Road, the vendors' income distribution



The histogram point out the distribution of incomes of the vendors. The large variety of order of magnitude calls for a logarithmical rescaling of the y-axis. The histogram points out that the distribution is extremely heavy-tailed due to the presence of some notable cases of success with respect to the sea of average incomes.

Figure 1.5: Silk Road, the vendors' income evolution curves



The evolution of incomes has been constructed smoothening the data over a two-week rolling time window. From the evolution of the curves, we can discriminate between the vendors that have constantly beaten the competition and those that had just a singulars successes. We are interested in understanding what are the attributes that permit such a steady and successful growth.

1.3.3 Silk Road customers analysis

ECLAT Basket Analysis

The plot in Figure 1.6 represents an application of the well know ECLAT algorithm for Association Rule Mining. Given a set of transactions, the algorithm finds rules that will predict the occurrence of an item based on the occurrences of other items in the transactions. In this specific case, the plot represents the Basket Analysis performed on the Silk Road's customers. The support is the number of times that a combination of products has been purchased over the total number of transactions.

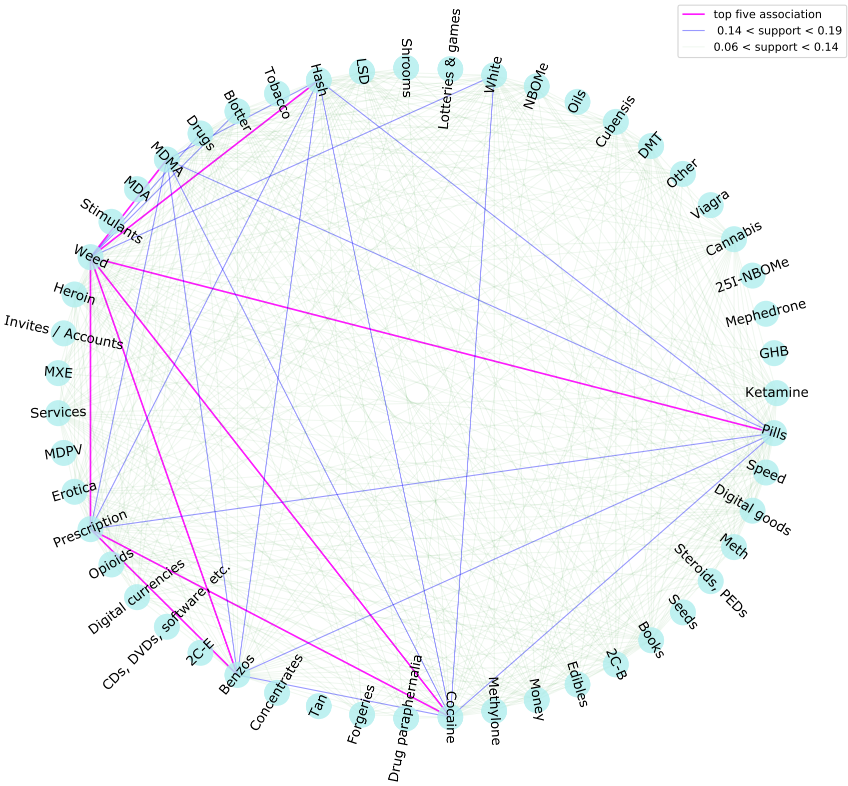
Analysis of the Biggest Buyer

Parsing the Silk Road dataset we found a buyer that dominated the market with 3428 total transactions. He interacted with 628 sellers over 141 different categories of products, across multiple countries. The main categories of its business, shown in Figure 1.7, are better explained in the following list:

- Category 22 = "Weed", with 429 transactions
- Category 122 = "Books", with 390 transactions
- Category 129 = "Erotica", with 163 transactions
- Category 63 = "Benzos", with 156 transactions
- Category 23 = "Hash", with 134 transactions

A visual analysis of Figure 1.7 shows that during the period of activity there has been a reduction in waiting times, a steady increase in the number of transactions and a clear growth in the basket diversification. As if the buyer had increasingly raised his trust in the

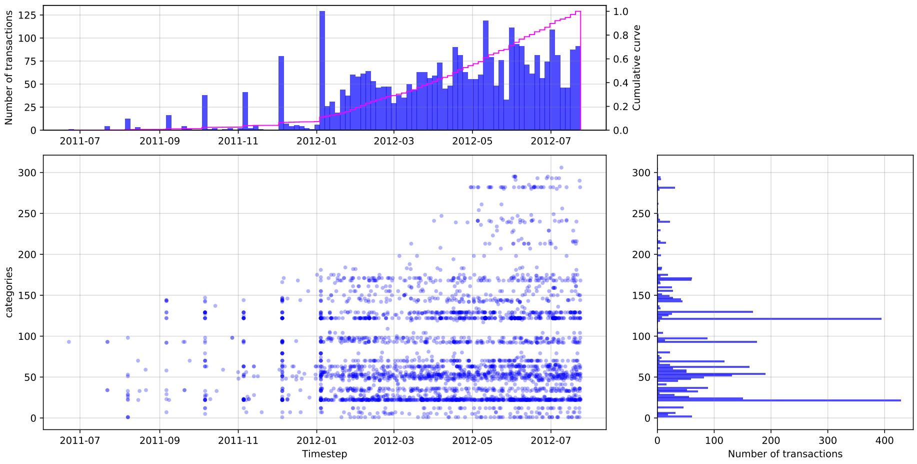
Figure 1.6: Silk Road ECLAT Basket Analysis, from the perspective of customers



The graph shows the habit of the customers in the Silk Road marketplace. In this case, we have only plotted the top fifty products of the market and we have just considered the combinations of a pair of products in the customers' baskets, in order to be able to display them on a graph. It is clear that *Weed* is commonly present in the basket of those who buy other illicit drugs, such as *MDMA*, *Cocaine*, *Pills* and *Benzos*. An interesting fact is that *Books* are able to move a notable amount of transactions in the market, as mentioned in Section 1.3.3, without being bought in strict combination with something else. Customers that buy *Books* tend to belong to a different category of customers, the Big Buyer of Section 1.3.3 seems to be an exception.

DNM. This is probably the reason why in Figure 1.8 we can appreciate the buyer eagerness in trying new sellers at each timestep.

Figure 1.7: Silk Road, scatter plot of the products' categories over time



Scatter plot of the products' categories over time. It clearly shows the growth in market diversification.

Dynamical approach of the Big Buyer shopping

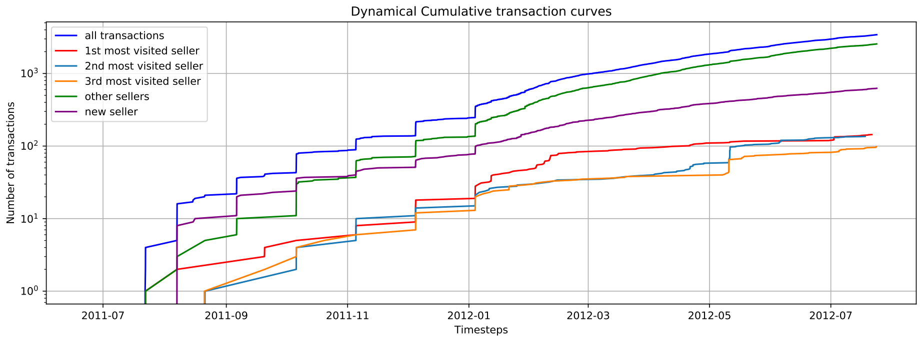
We are interested in understanding what kind of interactions the Big Buyer have instaurated with the sellers. The plot in Figure 1.8 picks a dynamic rank of seller popularity, which updates upon each transaction. Every time the Big Buyer makes a purchase, the ranks are modified and the seller that was picked for the transaction is increased in popularity. Therefore, we can see how much the rank of a seller influences the buyer's decision. From this analysis, we can appreciate how the high level of trust between the agents stimulates the willingness of trying new vendors. Nevertheless, it is also clear how this last process can coexist with a customer loyalty process. An interpretation connected to what has been seen in Figure 1.7, is that the Big Buyer is loyal to some vendors for certain products, while, as the market diversification grows, he is willing to try new products from new vendors.

1.4 A new relation between trust, competitiveness and evolution

From the review of the literature about the Darknet marketplaces, we understood that both the forum administration and the escrow system allowed buyers and sellers to transact without trusting each other since they both trusted the system. Nonetheless, results given by the exploratory data analysis suggested that customers still had preferences for some specifics vendors, no matter the anonymity and the ensured trustworthiness.

This forced us to revise our initial question on trust formation in high-risk environments, shifting our focus to a less-examined aspect of these peculiar markets, the establishment

Figure 1.8: Silk Road Big Buyer, the dynamical cumulative transactions curves



Dynamical Cumulative transaction curves of the Big Buyer plotted on a logarithmic scale. The dynamical categorization of the vendors is due to a ranking of how much they have been visited until that timestep. The jumps in the curves can be explained by an endogenous defect in the dataset, probably due to the scraping process made on specific windows of time.

of vendors' competitiveness. In the previous Section 1.3 we have shown that some vendors have constantly beaten the market competition even if the customers demand mainly concentrated on the most ubiquitous products. Then, why are these vendors more appreciated? What shape the vendor's competitiveness and what shape the sensitivity of buyers to vendors' attributes?

We conjectured that between the multitude of causes that enter in the matching process between vendors and customers there may be some hidden variables, such as the level of trustworthiness of each vendor or the degree of eagerness from the buyer. Ultimately, the complex range of dynamics that enter in this mechanism of competition pushed us to embrace an economic complexity approach, in which we only try to measure the unique set of intangible capabilities that make a vendor more attractive than another.

In this context, we are going to refer to this aggregate as the fitness of the vendor. Ultimately, the fitness is considered as the mathematical attribute that influences the customer to buy from a specific vendor, an inherent competitive factor that vendors may have and that is capable of affecting the network's evolution.

Chapter 2

Theoretical framework I

On the following, we present a digression on the economic complexity literature that will be of fundamental importance in order to understand how it aligns with Barney's organizational capabilities theory and how, this logical bridge, permits to scale down, to an individual level, arguments that have been originally applied on a national level.

We first recall the literature on strategic management and we relate it to the concept of economic complexity, subsequentially we briefly review the different algorithms that have been developed in the last years, pointing out the inconsistencies of the Hidalgo & Haussman's metrics, and the better choice of metrics proposed by Pietronero et al. and their Fitness-Complexity Algorithm. We then move to analyse the convergence behaviours of the algorithm and in which circumstances it encounters a loss in the cardinality of the fitness values; a feature of primary importance.

Finally, at the end of this section, we propose the generalization of the Fitness-Complexity Algorithm, in order to enlarge the analysis to a tripartite network of vendors, products and customers and extend even more the connection with Barney's model of sustained comparative advantage.

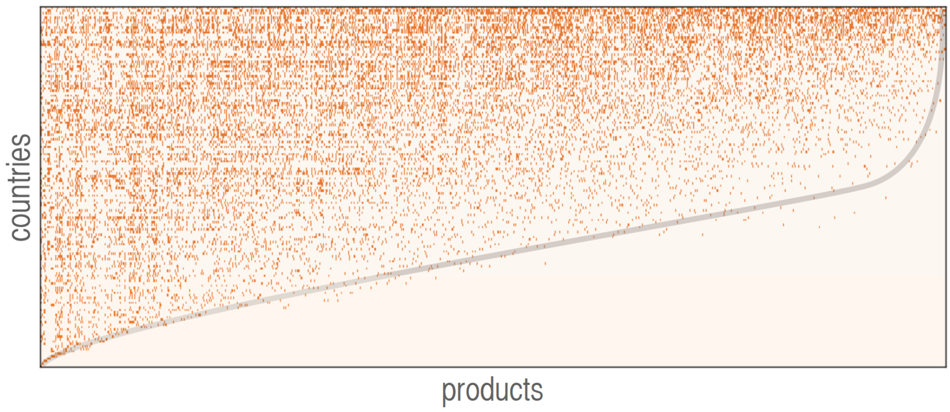
2.1 The concept of Economic Complexity

The prevalent macroeconomic analysis is based on a set of highly intangibles variables, such as good education, energy availability, labour cost, financial status, etc. Over the years, this process has led to a multitude of different theories and controversies, due to the subjective and informal process of weighing the single variables of the same reality. The goal of the Economic Complexity has been that of summarizing all these different analyses within a unique, impartial, non-monetary, non-income based and comparable metric. The concept of Economic Complexity try to address the problem of competitiveness and robustness of different countries in the global economy by studying the differences in the Gross Domestic Product and assuming that the development of a country is related to "different capabilities", considered as all the intangibles assets which drive the development, the wealth and the competitiveness of a country [9, 10]. While countries cannot directly trade capabilities, it is the specific combination of the latter that result in different products traded. More capabilities are supposed to bring higher returns and the accumulation of new capabilities provides an exponentially growing advantage. This approach is based on the idea that the productive basket of a country is able to discount and reflect all the information encoded in the intangible assets, usually hardly modelable. It resembles the so-called efficiency property of financial markets where, at least in principle, prices should reflect all the available sources and information, [9, 10, 11, 12, 13].

2.1.1 Diversification versus Specialization

The strongly dynamical context of the world trade economy suggests that flexibility and adaptability are the most important features of most competitive countries. This statement is clarified by a visual analysis of the Standard International Trade Classification data of the countries export flows, Figure 2.1. Visualizing the country-product matrix, sorting countries by diversification and products by the most to the less exported, it appears as a triangular matrix structure. The analysis reveals that there is a systematic relationship between the diversification of countries and the ubiquity of the products they make. More countries are complex and competitive more are diversified. The results are in contrast with traditional macroeconomics approaches, such as the Ricardian paradigm, which predict that the most successful countries should specialize in those products in which they score the highest comparative advantage. This approach would inevitably predict a block diagonal matrix.

Figure 2.1: The World Trade countries-products ordered binary matrix



An analysis of the Standard Trade Classification data of the export flows of different countries, assembled on a matrix. The Balassa's Revealed Comparative Advantage has been used in order the define if a country is able to compete with a specific product in the international market. The matrix is built from the export flows of the year 2010 and the products are categorized according to the Harmonized System 2007 coding system at 4 digits level of coarse-graining, [12].

2.1.2 Methodology

In order to define a suitable economic metrics to compare the trades of different countries in different products, the authors of [10, 12], used Balassa's Revealed Comparative Advantage (RCA). The Revealed Comparative Advantage is the fraction of product p in the export basket of the country c, with respect to the fraction of the world export of product p with respect to the total world export. For the sake of notation, q_{cp} represents the quantity of product p exported by country v.

$$RCA_{cp} = \frac{\frac{q_{cp}}{\sum_{p'} q_{cp'}}}{\frac{\sum_{c'} q_{c'p}}{\sum_{c'p'} q_{c'p'}}} \tag{2.1}$$

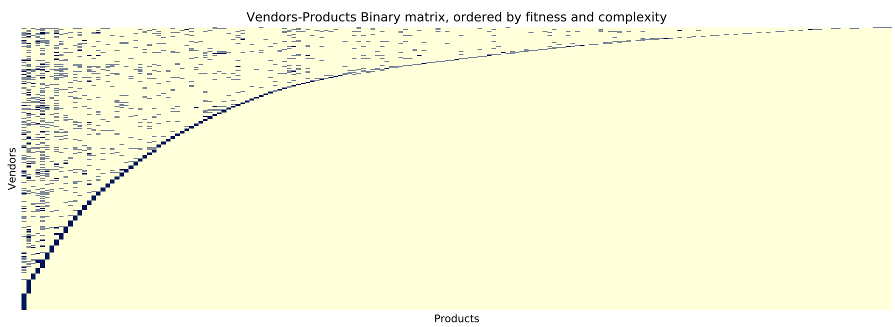
By adhering to the definition of RCA, they considered a country c to be a competitive exporter of a product p if the value RCA_{cp}, for such product, overcomes some minimal threshold RCA^*, $RCA^* = 1$. This procedure permits to build the binary elements, M_{cp}, of the $countries - products$ matrix. The elements equal to 1 state that country c has a comparative advantage in selling product p with respects to other countries, vice-versa for the elements equal to 0.

$$M_{vp} = \begin{cases} 1, & \text{if } RCA_{vp} > RCA^* = 1 \\ 0, & \text{otherwise} \end{cases} \tag{2.2}$$

2.1.3 Barneys' model as a scale-down argument

Hidalgo & Haussman's theory of capabilities bases its assumption exactly on the nested structure of the binary matrix of countries and products and on the competitive advantage given by diversification over products. In our case, the Silk Road bipartite network of vendors and products shows the same triangular structure in the binary matrix and, thus, permits us to make the same conceptual statement, Figure 2.2.

Figure 2.2: The Silk Road vendors-products ordered binary matrix



The ordered binary matrix of the Silk Road Darknet market. The dataset is related to the total market activity and the Balassa's Revealed Comparative Advantage has been used in order the define if a vendor is able to compete with a specific product in the Silk Road Darknet market.

As previously introduced, we conjecture that the relation between diversification and competitive advantage, expressed by the nested structure of the matrix, is logically connected with the principles of strategic management developed in [8]. The distribution of products by complexity, their stability over time and their relation with the future development of a vendor it's easily relatable with the four empirical principles of sustained comparative advantage mentioned in Barney's model. Value, rareness, imperfect imitability and substitutability. Again, this logical bridge is of primary relevance because it permits us to confidently scale down arguments successfully applied on a national level to an individual/organizational level. Practically, moving from the global market to Darknet markets, we can refer to the previous methodological Section 2.1.2, substituting countries c with vendors v.

2.2 Economic Complexity Index vs Fitness-Complexity Algorithm

2.2.1 Hidalgo-Haussman metrics and the Economic Complexity Index

In this spirit, the first attempt of defining the competitiveness in terms of diversification has been made by Hidalgo & Haussmann, [10], who came out with an iterative algorithm that leads to a measure of the Economic Complexity Index. Their fundamental idea is that the complex set of capabilities of a country determines the nested structure of the country-product matrix. With the significance that ubiquitous products require few capabilities and therefore they can be produced by most of the countries, while diversified countries are the only one to have all the necessary capabilities required for the most complex and rare products. They proceeded to define the fitness, F, of a country and the ubiquity Q, of a product in an iterative way, the so-called 'Reflections method'. (We have used the Pietronero et al. notation, for the sake of convenience and consistency).

$$\begin{cases} F_c^{(n)} = \dfrac{1}{k_c} \displaystyle\sum_{p=1}^{N_p} M_{cp} Q_p^{(n-1)} & (2.3) \\[2ex] Q_p^{(n)} = \dfrac{1}{k_p} \displaystyle\sum_{c=1}^{N_c} M_{cp} F_c^{(n-1)} & (2.4) \end{cases}$$

At the 0^{th} order the fitness of a country is given by the number of competitive products produced, which is called diversification of such country, k_c. Similarly, the ubiquity of a product is the number of the countries producing such a product, the degree k_p.

$$\begin{cases} F_c^{(0)} = \displaystyle\sum_{p=1}^{N_p} M_{cp} \equiv k_c & (2.5) \\[2ex] Q_c^{(0)} = \displaystyle\sum_{c=1}^{N_c} M_{cp} \equiv k_p & (2.6) \end{cases}$$

At the 1^{st} order, these iterated fitnesses and ubiquities have simple interpretations: the fitness $F_c^{(1)}$ is the average ubiquity of the competitive products produced by a country c. Similarly, the ubiquity $Q_c^{(1)}$ is the average fitness of the countries producing the product p. Larger order interpretations become less and less intuitive. It has been proven that the convergence of the algorithm is independent of the initial conditions.

2.2.2 Pietronero et al. metrics and the Fitness-Complexity Algorithm

In a sequence of different papers, [11, 12, 13, 14, 15], Pietronero et al. pointed out the illogical inconsistencies of Hidalgo & Hausmann's metrics and proposed new ones in the so-called Economic Fitness model. Nevertheless, the authors share the same assumptions on the theory of capabilities and use the same nested structure of the country-product matrix as their analysis starting point. They outlined that the conceptual problems of the Economic Complexity Index Algorithm, are of two natures:

- The averaging process cuts out the importance of the diversification, producing unrealistic results. For example countries with 10 products of complexity that goes from 1 to 10 have lower fitness, 5.5, that countries with a single product of fitness 6.

- The linearity of the relation between the complexity of a product and the average fitness of a country that is able to produce it leads to fundamental inconsistencies. For example, consider Crude Oil and in general any raw material of simple extraction. Their availability is just a matter of luck and it is not a reflection of the industrial competitiveness of a country. Despite this, the Hidalgo-Hausmann metrics push up the complexity of crude oil due to the presence of high fitness exporters such as Norway, USA, UK.

To overcome these inconsistencies Pietronero et al. proposed a new metrics for the country fitness and for the product complexity, taking inspiration from the Google PageRank algorithm [11]. They clearly stated the importance of having an algorithm that explicitly includes the diversification as an essential concept. Thus, they decided to still frame the fitness of a country, F_c, into a linear sum over all the competitive products weighted by the complexity, Q_p, of each product.

For the complexity of a product, Q_p, Pietronero et al. brought a more subtle and highly non-linear definition. They clearly wanted to introduce the fact that, in general, the more are the countries that can make a product, the less that product is complex. Moreover, they have included the fitness of each country, because each of them must enter with a different weight in the definition of the complexity of a product. For example, if we consider that Germany, USA and China are the only (nowadays high fitness) countries that produce a certain product, a linear metric would correctly describe this product as a highly complex one. However, if we consider another product, exported also by Nigeria (nowadays a low fitness country), the same metric would decrease the complexity of this product linearly. The comparison between the complexity of the two product wrongly reflects the fact that Nigeria is actually able to produce only this and a few other products, implying that the product that Nigeria shares with the other three countries should require just a really small set of capabilities and thus its complexity should decrease way more. These considerations drove Pietronero et al. to propose a non-linear metric for the complexity of a product, in which the country with the lowest fitness, that can make the product, give the maximum handicap to the value of the complexity of this product [11, 12, 13, 14, 15]. Figure 2.3 shows an example of the non-linear correction.

Ultimately, it is necessary to add the normalization equations, equations 2.9 and 2.10, in order to escape the two trivial absorbing solutions of 0 and $+\infty$:

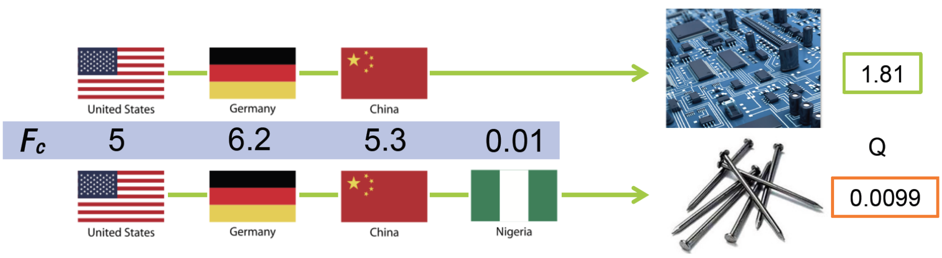
Step 1.

$$\begin{cases} \tilde{Q}_p^{(n)} = \left(\dfrac{1}{\sum_{v=1}^{N_c} M_{cp} \dfrac{1}{F_c^{(n-1)}}} \right) & (2.7) \\ \tilde{F}_c^{(n)} = \displaystyle\sum_{p=1}^{N_p} M_{cp} Q_p^{(n-1)} & (2.8) \end{cases}$$

Step 2.

$$\begin{cases} F_c^{(n)} = \dfrac{\tilde{F_c}^{(n)}}{\langle \tilde{F_c}^{(n)} \rangle_c} & (2.9) \\ Q_p^{(n)} = \dfrac{\tilde{Q_p}^{(n)}}{\langle \tilde{Q_p}^{(n)} \rangle_p} & (2.10) \end{cases}$$

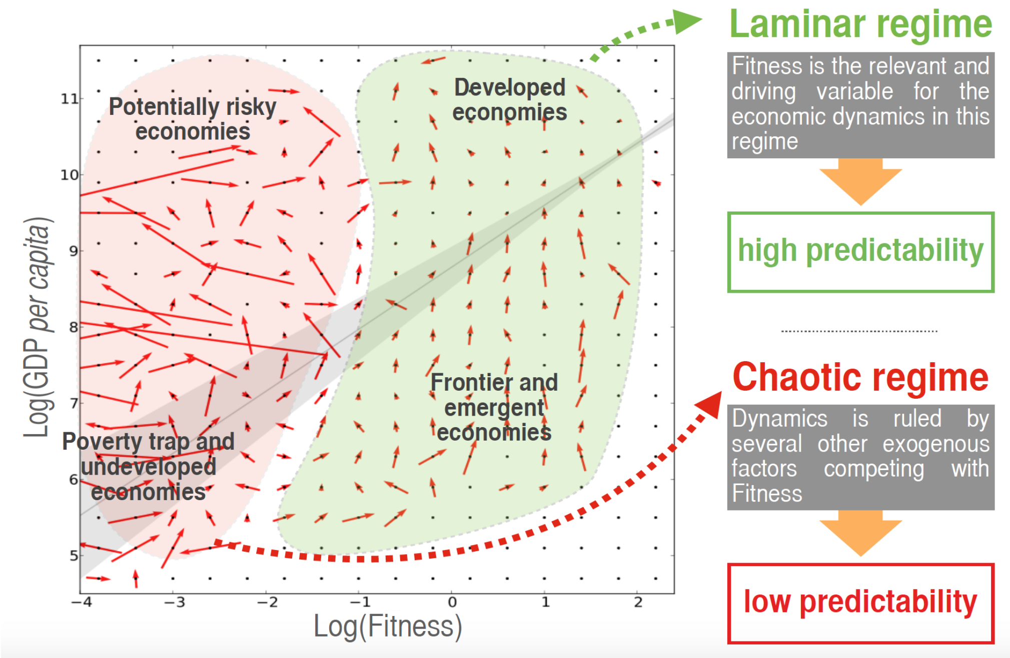
Figure 2.3: The extreme non-linearity of the products' complexity



The extreme non-linearity of the product complexity implies an extreme dependence on the lowest fitness producer, [12].

These new metric give surprisingly good results. The study of the countries time trajectories in the GDP-Fitness space has been successfully utilized to forecast the development of countries over the years, the idea is that the comparison of this intrinsic competitiveness with the GDP can reveal the hidden potential for the development of a country. In Figure 2.4 we can see a flowchart constructed from the coarse-grained time trajectories in the GDP-Fitness space, of all countries, for the past 15 years [13]. It shows an interesting

Figure 2.4: Countries'dynamics in the Log GDP - Log Fitness space



A finer coarse-graining of the dynamics highlights two regimes for the dynamics of the evolution of countries in the fitness-income plane. There exists a laminar region in which fitness is the driving force of the growth and the only relevant economic variable in order to characterize the dynamics of countries. There is also a second regime, which appears to be chaotic and characterized by a low level of predictability, [13].

heterogeneous structure, in the region of high fitness (green), we observe a smooth, laminar behaviour of the flow, while in the low fitness region (red) we have a rather irregular and chaotic dynamics. In the laminar-like regime, the fitness is the relevant economic variable that permits to understand the dynamics of the income and, in general, of the growth of the GDP. In the other regime, instead, the dynamics is ruled by several exogenous factors which compete with the fitness in driving the evolution of countries. It results that the predictive scheme, required by the dynamics of economic complexity, is analogous to the problem of predicting the future of a dynamical system in the case in which we do not know the equations of motion.

In Chapter 3, we discuss how we have tried to replicate this dynamical approach to our network of vendors, products and customers.

2.3 On the convergence of the Fitness-Complexity Algorithm

Having said that, we underline that all the cardinal and ordinal values of the countries' fitnesses are determined only by the shape of the countries-products binary Matrix, M. In principles, obviously, it is possible to apply the Algorithm to any bipartite network defined by the binary matrix M, and, in our case, we will have vendors and products in place of countries and products, and Fitness of a vendor, F_v, and Fitness of a product, F_p, in place of Fitness of a country, F_c and Complexity of a product Q_p. On the following, we will also utilize a different notation for the elements of the binary matrix, M_{vp}.

2.3.1 The Matrix Class

The analysis of the Fitness-Complexity Algorithm follows step by step the paper "On the convergence of the Fitness Algorithm", a short note suggested by the authors of the algorithm that helped us understand the convergence behaviour on our different datasets. [16], [17].

The importance of convergence to non zero values

The paper analyses the characteristic of the matrix M required to have convergence to non zero values for the fitness of the vendor F_v and the fitness of the product F_p. It is substantial to underline that the convergence of the Fitness-Complexity Algorithm, to values strictly greater than zero, is required to maintain the cardinality of the fitness values, a feature of primary importance in order to quantitatively exploits their values in further analysis, retaining a fixed and a well-defined difference between them. In the case of a loss of cardinality, which may happen when some values exponentially decay to zero, we would be able only to retain the ordinal information given by the ranking of the different rates of decay; hence, the informativeness of the quantitative values would result deprecated by the decay itself. In fact, in such a case, it would be impossible to retrieve fixed and well-defined fitness values as they reach it (zero) in an infinite time.

The general Matrix Class considered in the paper, [16], is composed of 4 blocks, one of which of zeros. The other three blocks are a mixture of zeros and ones. On the following, the density of ones in the three blocks doesn't play any role in the convergence behaviour, thus, for the sake of simplicity, we can consider any kind of density. An example of the class is presented in Figure 2.5.

Figure 2.5: The general Matrix Class of the Fitness-Complexity Algorithm

$$
\begin{array}{cc}
 & \begin{array}{ccc|cccccc} & C_1 & & & & & C_2 & & \end{array} \\
\begin{array}{r} R_2 \left\{ \begin{array}{c} \\ \\ \\ \\ \end{array} \right. \\ R_1 \left\{ \begin{array}{c} \\ \\ \\ \end{array} \right. \end{array} &
\left(\begin{array}{ccc|cccccc}
1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
\hline
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{array} \right)
\end{array}
$$

The general Matrix Class of the Fitness-Complexity Algorithm, [16].

More importantly, the authors of the paper state the following ansatzes, [16]:

Ansatz 1 If the belly of the matrix is outward with respect to diagonal, all the fitnesses and complexities converge to numbers greater than 0. If the belly of the matrix is inward, some of the fitnesses will converge to 0. The belly of the matrix is given by the profile of the external 1s, those that confine to the region populated by zeros, the so-called external region. The belly is considered to be $inward - belly$ if the diagonal of the matrix cross the external region, while it is considered to be $outward - belly$ if the diagonal of the matrix cross the profile of external 1s.

Ansatz 2 A matrix M have all the fitness and complexity different from zero if, after ordering said matrix, the diagonal line does not pass through the external area.

Ansatz 3 If the diagonal line does pass through the external part of an ordered matrix M, some countries and products will converge to zero. If we progressively remove them from the analysis, defining new ordered matrices M, we will have a convergence to finite values of fitnesses and complexities for the remaining countries and products for which the new M has a diagonal line not passing through the external area.

This last ansatz defines the $crossing - country$ point of a matrix as the lowest fitness country that converges to a non zero fitness, in our case, we will call it $crossing - vendor$ point. An example of it is shown in Figure 2.6. More in detail, the procedure to obtain a remaining matrix with all the vendors converging to non zero values, is the following:

Crossing country procedure

1. Run the Fitness algorithm on the total dataset and obtain the so-called $ordered - matrix$, the matrix M obtained after the rearrangement of the columns and rows such that all the countries and all the products are ranked accordingly to their fitness.

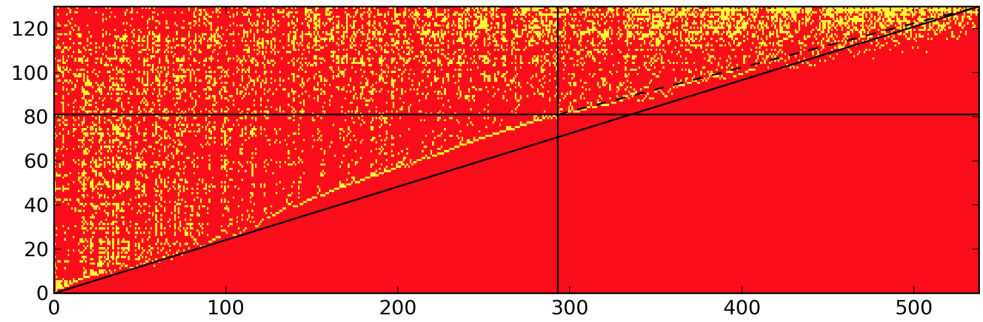
2. Remove the least fit country and all the products that have a complexity minor or equal to the most complex product sold by the country.

3. Go to step 1 and break when all the vendors' fitnesses are converging to non zero values.

The oligopolies' effect and the exponential decay

In the paper [16], the authors state clearly the importance of oligopolies and, more generally, N-poly, for the convergence behaviour. In the case in which one or more N-poly has a further N-poly which is not shared, the decay of the competitor will be exponential. We can look at Figure 2.7 as an example. The first country has two monopolies, while country 3 one monopoly and countries 2 and 4 no monopolies at all. All countries tend to zero exponentially faster, except for the first one. In terms of the previous ansatzes, matrices that are decaying exponentially are always inward belly, exactly as the matrix in Figure 2.7. Moreover, if we run the $crossing - country$ procedure we would end up with the trivial 1x1 matrix composed only by the first country. Basically, this is why the algorithm treats all the other countries differently, making them converge to zero exponentially.

Figure 2.6: The binary matrix's crossing country point



The matrix frontier is consistently above the diagonal line. The horizontal and vertical lines represent the number of countries and products that converges to zero. The dashed line shows the diagonal of the new block remaining after removing those countries and products: the diagonal of the new M matrix now does not cross the external part of the matrix, [16].

Unfortunately, we will see that our class of matrices belong exactly to this particular case. The presence of vendors selling really specific products creates a network in which we have a large presence of oligopolies and monopolies. Along the Silk Road and Agora timelines, vendors diversify their listings in the most various ways. By the introduction of new products, they exasperate the $inward - belliness$ of the matrix and the number of vendors' fitnesses that decay exponentially.

Figure 2.7: The exponential decay

$$\begin{array}{c} c \\ e^{-n} \\ e^{-n} \\ e^{-n} \end{array} \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The Matrix Class with an exponential decay, [16].

Possible solution in case of loss of cardinality

In general, it seems that in case a significant fraction of countries-products goes to zero, it is common to adopt one of the following strategies, [17]:

1. We could make use of the values' ranking instead of their values. We should still obtain reasonable results without impoverishing the dataset.

2. We could look for a $crossing - vendor$ point. A strategy that that in our case seems to not be feasible, due to the exaggerated $inward - belliness$ and consequential convergence of the $crossing - country$ procedure to the trivial 1x1 matrix. Anyway, in Section 2.2.2, we will treat some examples of constrained datasets.

3. We could abruptly stop the algorithm at a specific iteration and use those quantitative values of fitnesses. Even this $less - educated$ option would not give reliable results in terms of cardinality, as, after few iterations, most of the vendors' fitnesses are already approximately zero.

In Chapter 3, after trying all the different strategies, we finally opted for just continue with the ranking values of the fitnesses.

2.4 Tripartite network and metrics' generalization

We have seen that the Fitness-Complexity Algorithm can lead to controversial results in case of exponential decays, and as previously mentioned, if applied on our bipartite case of vendors and products it suffers from the same problem.

Let's recall that the goal of this Master's Thesis is to successfully analyze the evolution of the Darknet markets and the establishments of competitiveness between vendors. In order to do so, we proposed to tackle such questions into the framework of economic complexity, building a parallel with the global market. However, in order to proceed with this coherently treat such economic network, we should also consider how to insert in the model the extra presence of customers. This is, in fact, an unavoidable difference with respect to the bipartite network of the global economy. On the contrary of countries that trade products with each other, vendors sell their products to a totally different class of agents, customers. Thus, it has seemed necessary for such kind of market, to target the competitiveness of a product also by its ability to impact the market demand and not only by its degree of ubiquity.

In view of that, we have detected two different ways of generalizing the metrics to a tripartite network of vendors, customers and products:

1. We can reduce the number of customers to those that have a real impact on the market, for example, those that have bought more than ten times. This reduction makes the process computationally treatable and permits us to consider each customer differently. In this framework, we can try to generalize the metrics of Pietronero et al. by adding a new variable, related to the fitness of customers, F_c.

2. We can treat all the customers equally, as a sort of external noise of the market. Furthermore, we can consider the noise capable of leaving informative traces that shape the market dynamics. In this framework, we can exploit the information that customers leave after their transactions, such as the feedback rating of the vendors.

We have summarized all the proposed generalisations in the Table 2.1.

2.4.1 Case 1, customers treated as autonomus agent

As previously discussed, we can try to generalize the system of equations by introducing customers as singular agents. This procedure requires to introduce their fitnesses, F_c and their binary matrices, the vendors-customers one, M_{vc}, and the customers-products one, M_{cp}. In the formulation of the new metrics we expect that the fitness of a customer describes its impact on the market, namely, it should define how the customer transactions boost the market activity.

Our new approach is defined in the following way:

- In the spirit of Pietronero et al metrics, we suggested defining the fitness of a product, F_p, in the same way they defined the fitness of countries exports. Inversely proportional to the ubiquity of the product in the vendors' catalogues, M_{vp}, and highly proportional to the lowest fitness vendor, F_v.

- We define the fitness of a vendor, F_v, proportional to the number of different customers that he reaches, M_{vc}, weighted by their fitness F_c. A high-fitness vendor is a person who sells to many different high-fitness customers.

- Similarly, we define the fitness of a customer, F_c, proportional to the number of different products that he buys, M_{cp}, weighted by their fitness, F_p. A high-fitness customer is a person who buys many different high-fitness products.

Step 1:

$$
\begin{cases}
\tilde{F}_p^{(n)} = \left(\dfrac{1}{\sum_{v=1}^{N_v} M_{vp} \dfrac{1}{F_v^{(n-1)}}} \right) & (2.11) \\[2ex]
\tilde{F}_v^{(n)} = \displaystyle\sum_{c=1}^{N_c} M_{vc} F_c^{(n-1)} & (2.12) \\[2ex]
\tilde{F}_c^{(n)} = \displaystyle\sum_{p=1}^{N_p} M_{cp} F_p^{(n-1)} & (2.13)
\end{cases}
$$

Step 2:

$$
\begin{cases}
F_p^{(n)} = \dfrac{\tilde{F_p}^{(n)}}{\langle \tilde{F_p}^{(n)} \rangle_p} & (2.14) \\[2ex]
F_v^{(n)} = \dfrac{\tilde{F_v}^{(n)}}{\langle \tilde{F_v}^{(n)} \rangle_v} & (2.15) \\[2ex]
F_c^{(n)} = \dfrac{\tilde{F_c}^{(n)}}{\langle \tilde{F_c}^{(n)} \rangle_c} & (2.16)
\end{cases}
$$

2.4.2 Case 2, customers treated as an informative noise feedback

We keep the definition of the Fitness-Complexity Algorithm. Products are considered to be competitive, high F_p, if they are rare and produced by competitive vendors. Whereas, vendors are considered to be competitive, high F_v, if they are largely diversified, capable of selling high fitness products and appreciated by customers. The customers' appreciation is obtained by the feedback rates that they need to release after each transaction. Feedback ratings, here called R_{vp}, are usually related to the appreciation of the product's quality and the quality of the vendor's service [2, 4, 5, 7]. It is known, that the rating of a vendor, with respects to his products and his service, is a property that conserves the memory over the timesteps.

The introduction of equation 2.19 permits the update, at time t, of the mean of the vendor-product rating, $\tilde{R}_{vp}^{(t)}$, with respect to its mean at time t-1. This process permits to keep the memory of ratings. Therefore, the set of the equations is:

Step 1:

$$\begin{cases} \tilde{F}_p^{(n)} = \left(\dfrac{1}{\sum_{v=1}^{N_v} M_{vp} \dfrac{1}{F_v^{(n-1)}}} \right) & (2.17) \\[2ex] \tilde{F}_v^{(n)} = \displaystyle\sum_{p=1}^{N_p} M_{vp} F_p^{(n-1)} \tilde{R}_{vp}^{(t)} & (2.18) \\[2ex] \tilde{R}_v^{(t)} = \displaystyle\sum_{p=1}^{N_p} R_{vp}^{(t-1)} + \frac{1}{t} (R_{vp}^{(t)} - \tilde{R}_{vp}^{(t-1)}) & (2.19) \end{cases}$$

Step 2:

$$\begin{cases} \hat{F}_v^{(n)} = \dfrac{\tilde{F_v}^{(n)}}{\langle \tilde{F_v}^{(n)} \rangle_v} + \dfrac{\tilde{R_v}^{(n)}}{\langle \tilde{R_v}^{(n)} \rangle_v} & (2.20) \\[2ex] F_p^{(n)} = \dfrac{\tilde{F_p}^{(n)}}{\langle \tilde{F_p}^{(n)} \rangle_p} & (2.21) \end{cases}$$

Step 3:

$$\begin{cases} F_v^{(n)} = \dfrac{\hat{F}_v^{(n)}}{\langle \hat{F}_v^{(n)} \rangle_v} & (2.22) \end{cases}$$

The third normalization step, 2.22, is required to keep the average values of the fitness equal to 1, in order to escape the two trivial absorbing solutions of 0 and $+\infty$.
We discuss the results of these generalizations, 2.4.1 and 2.4.2, in Chapter 4.

Table 2.1: Metrics' generalizations on the tripartite network

Generalized equations	Main variation	Results
$\tilde{F}_p^{(n)} = \left(\frac{1}{\sum_{v=1}^{N_v} M_{vp} \frac{1}{F_v^{(n-1)}}} \right)$ $\tilde{F}_v^{(n)} = \sum_{c=1}^{N_c} M_{vc} F_c^{(n-1)}$ $\tilde{F}_c^{(n)} = \sum_{p=1}^{N_p} M_{cp} F_p^{(n-1)}$	Customers treated as autonomous agents, with their own fitness, $F_c^{(n)}$, proportional to the vendor's fitness, $F_v^{(n-1)}$	Results shown in Chapter 4
$\tilde{F}_p^{(n)} = \left(\frac{1}{\sum_{v=1}^{N_v} M_{vp} \frac{1}{F_v^{(n-1)}}} \right)$ $\tilde{F}_v^{(n)} = \sum_{p=1}^{N_p} M_{vp} F_p^{(n-1)} \tilde{R}_{vp}^{(t)}$ $\tilde{R}_v^{(t)} = \sum_{p=1}^{N_p} R_{vp}^{(t-1)} + \frac{1}{t}(R_{vp}^{(t)} - \tilde{R}_{vp}^{(t-1)})$	Customers included as an external noise. We have considered their feedback raiting, $R_{vp}^{(t)}$.	Results shown in Chapter 4

Listings of all the proposed generalizations. For the sake of a lighter notation, on all the generalizations we have omitted the equations related to the normalizations process. Step 2 for 2.4.1, Step 2 and Step 3 for 2.4.2.

Chapter 3

Bipartite network's results

3.1 Results on Silk Road

3.1.1 The convergence behaviour of the total market activity

The first step has been to implement the Fitness-Complexity Algorithm and to fit it with the Silk Road dataset in order to be able to build the binary matrix, properly ordered by vendors' fitnesses and products' fitnesses. The shape of the binary matrix gave us an idea of the possible convergence behaviour of the fitness values, and consequentially, it helped us interpreting the reliability of the quantitative results. Initially, we have considered the total market activity, while later, we have experimented with some constrained Silk Road's dataset.

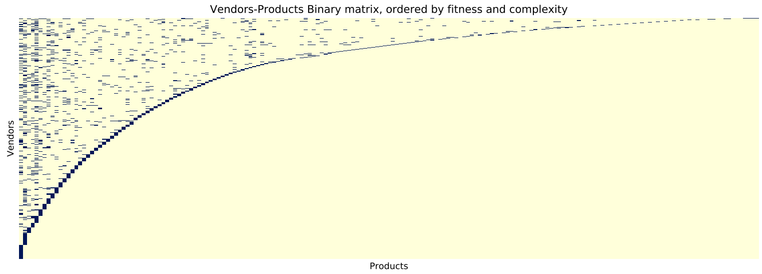
The total market activity of the Silk Road dataset is composed of:

- all the vendors that have earned an amount different from zero, for a total number of 989 vendors.

- all the product categories that have moved an amount of money different from zero, for a total number of 187 categories.

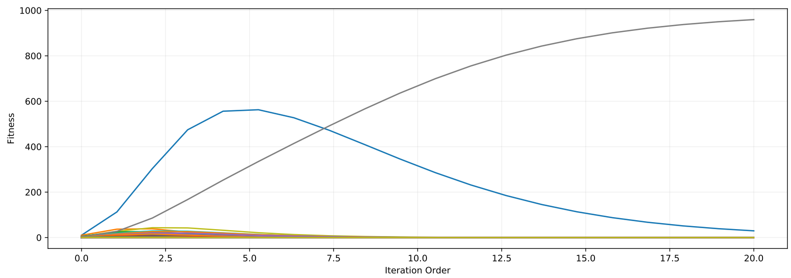
The shape of the matrix is shown in Figure 3.1, and, as previously mentioned, it clearly belongs to the exponential decay class. The ansatz 3, introduced in Section 2.3.1, would suggest us to implement a $crossing-country$ procedure in order to find out the $crossing-vendor$ point, [16]. Unfortunately, in our case, the only result is the trivial 1x1 matrix that contains only the fittest vendor and its fittest product. In fact, the $inward-belliness$ of the matrix M is so pronounced that it excludes any possibility of having an $outward-belly$ submatrix. As a consequence, only the first vendor and the first product converge to a non zero value while the other vendors and products decay exponentially, as clearly shown in Figure 3.2 and Figure 3.3.

Figure 3.1: Silk Road, the ordered binary matrix.



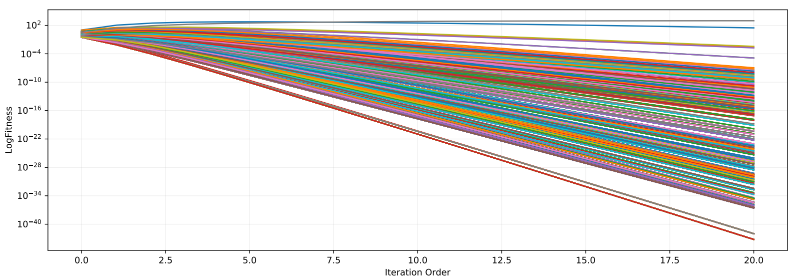
The ordered binary matrix of the Silk Road Darknet market. The dataset is related to the total market activity.

Figure 3.2: Silk Road, convergence behaviour of the vendors' fitnesses



Convergence behaviour of the vendors' fitnesses. The dataset is related to the total market activity of the Silk Road Darknet market.

Figure 3.3: Silk Road, logarithmic convergence behaviour of the vendors' fitnesses.



Logarithmic convergence behaviour of the vendors' fitnesses. The dataset is related to the total market activity of the Silk Road Darknet market.

3.1.2 The convergence behaviour of a constrained market activity

Nevertheless, for the sake of curiosity, we have also analyzed a constrained Silk Road datasets, related to the top twenty exported products in term of generated cash flow. This constraining drastically reduces the possibility of diversification between vendors and only considers the products that have really had an impact in the market and over the customer demands. It still permits us to consider the 85 % of the vendors and the 97.8 % of the market GDP.

The constrained ordered binary matrix and the new convergence of the vendors' fintess are orderly shown in Figure 3.4,Figure 3.5, Figure 3.6.

Even in this case, the number of vendors fitnesses that are converging to zero it is still a large fraction of the total, as predicted by the failure of the *crossin country procedure*, described in Section 2.3.1.

3.1.3 Vendors' dynamics of the total market activity

As we are interested in how the fitness, F_v, is able to describe the competitiveness of a vendor, we would like to compare the fitness with the income of each vendor in each specific time window. As previously stated, due to the lack of cardinality, we can only use the ranking of the vendors' fitnesses, an option that can still reveal insights on the informativeness of the metrics, according to [17].

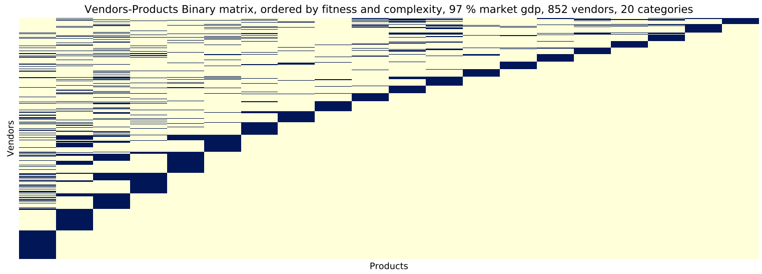
As a first step, we have implemented a rolling window over the total market activity of Silk Road, in order to smooth and accomplish the lack of data. In each time window, we have calculated the vector of vendors' fitnesses and the vector of vendors' incomes. Because of the entrance of new vendors in the market, the size of the vectors change along the timeline, and, in order to make a proper comparison between the ranking values at different times, it has been necessary to normalize their values between 0 and 1. For the sake of clarification, we specify that the 0th position is related to less fit vendor while the 1st position to the fittest one.

In Figure 3.7 we can appreciate how each vendor has evolved in terms of its Fitness rank. It shows how the richest vendors on the overall market activity are not those that have the highest fitness. We can already appreciate the first unexpected behaviour of the vendors' fitness, it is not true that the more you are fit the more you earn. This is just one of many insights that we will make us state the ineffectiveness of the Fitness-Complexity Algorithm as a growth indicator in the Darknet marketplaces.

In the spirits of the analysis of Pietronero et al, [13], we have also plotted the time trajectories of each vendor in the GDP - fitness space, Figure 3.8. Accordingly, to what has been shown in Figure 3.7, it is again evident that the there is not a clear correlation between high fitness and high incomes.

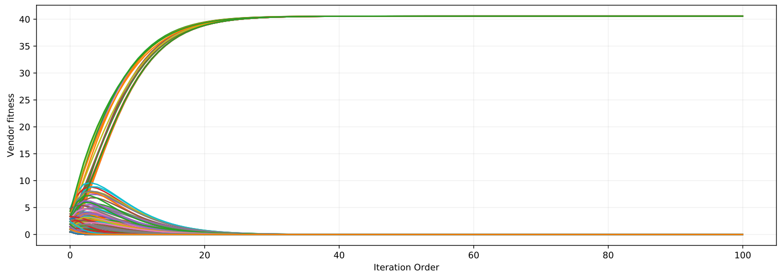
This results can be more easily appreciated in a coarse-grained version of the trajectories space. In Figure 3.9 we have defined a specific binning for the trajectories space and in each box, we have calculated the average incremental vector. The intention of the coarse-graining representation is that of representing the mean field movement in certain regions of the GDP - fitness rank space. In order to discount the average increase in cash flow in the market, we have subtracted the average income of all the vendors in each time window.

Figure 3.4: Constrained Silk Road (97.8 % GDP), the ordered binary matrix



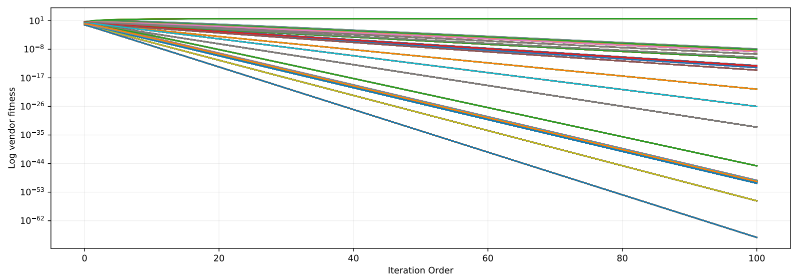
The ordered binary matrix of the Silk Road Darknet market. The dataset is related to a constrained market activity of Silk Road (97.8 % GDP), that takes into account the top 20 proficuous products.

Figure 3.5: Constrained Silk Road (97.8 % GDP), convergence behaviour of the vendors' fitnesses



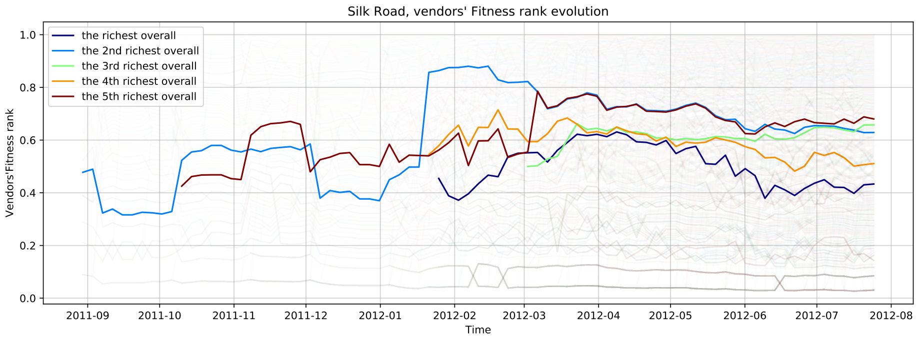
Convergence behaviour of the vendors' fitnesses. The dataset is related to a constrained market activity of Silk Road (97.8 % GDP), that takes into account the top 20 proficuous products.

Figure 3.6: Constrained Silk Road (97.8 % GDP), logarithmic convergence behaviour of the vendors' fitnesses



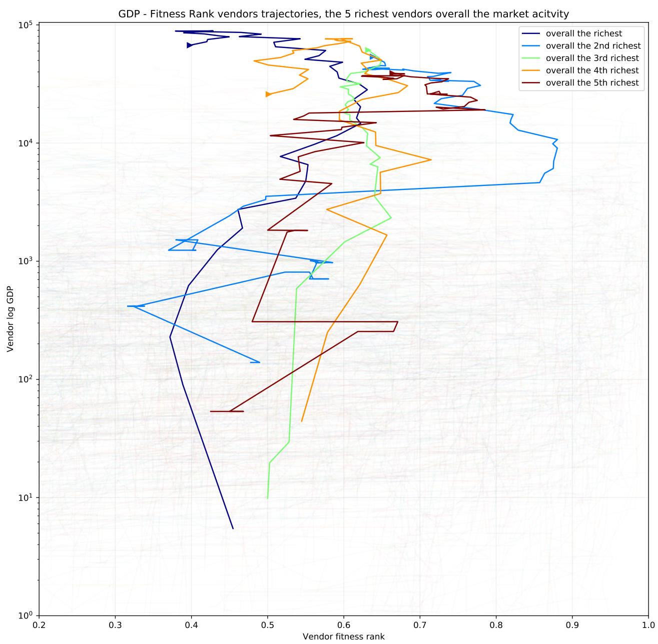
Logarithmic convergence behaviour of the vendors' fitnesses. The dataset is related to a constrained market activity of Silk Road (97.8 % GDP) that takes into account the top 20 proficuous products.

Figure 3.7: Silk Road, evolution of the vendor's fitness ranking



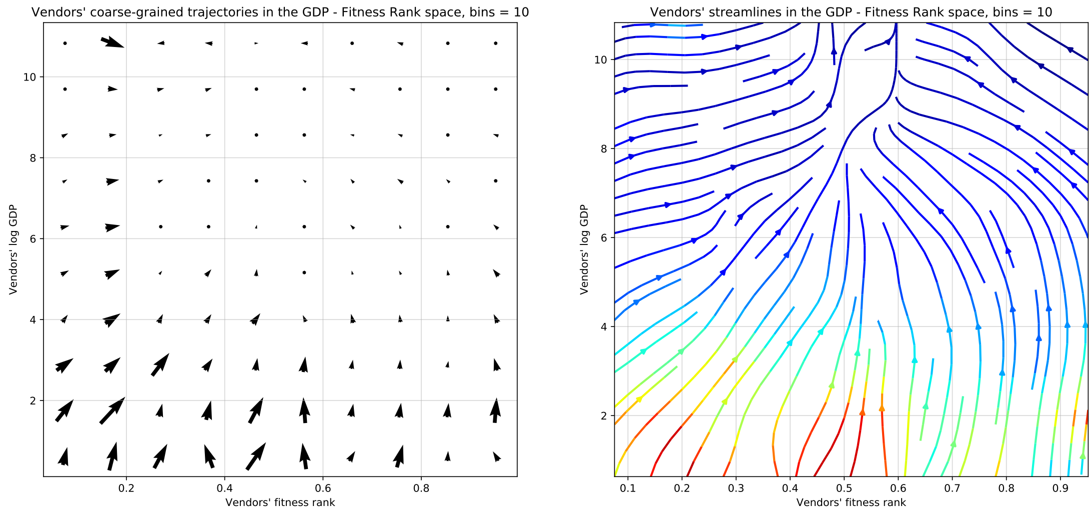
Evolution of the ranking of the vendors' fitnesses in the Silk Road Darknet market. The dataset is related to the total market activity. Of all the 989 vendors we have highlighted the ten richest vendors over the total timeline of the market activity. The entrance of new vendors is clearly visible in the figure.

Figure 3.8: Silk Road, Vendors' trajectories in the GDP - Fitness Rank space



Vendors' trajectories in the GDP - Fitness Rank space. The dataset is related to the total market activity of the Silk Road Darknet market. Of all the 989 vendors we have highlighted the five richest vendor over the total timeline of the market activity.

Figure 3.9: Silk Road, the arrow-plot and the streamlines of the vendors' trajectories in the GDP-Fitness space



The plot on the left is the arrows plot of the coarse-grained version of the vendors' trajectories in the GDP-Fitness space. While the plot on the right represents the streamlines of the vendors' trajectories in the GDP-Fitness space. The size of the arrows and colours state the norm of the incremental vectors in each bin, with an increasing value going from blue to red. We have considered only the boxes that contained more than 3 points. The dataset is related to the total market activity of Silk Road.

Moreover, in order to discard outliers behaviours, we have considered only the boxes that contained more than 3 points. The Figure 3.9 let us think about a basin of attraction in the dynamics of the vendors.

We can appreciate mainly three different regimes:

1. Vendors in a condition of a low income tend to increase it moving to the average income of the market. This trend can be explained by the fact that customers were always willing to try new vendors due to the high trustworthiness in the market, [3].

2. Vendors with a lower fitness and higher income tend to increase their fitness, we conjecture that this regime can be explained in different ways:

   - Vendors tend to reinvest their money in their activity, increasing their diversification of products and consequently their fitness.
   - Customers prefer vendors that have already been discussed in the forum, namely, those ones that have already carried out a substantial number of transactions and have already an high income. This interest of the customers can be the trigger for new transactions in new products category, in other words, customers are able to increase the diversification of the vendors of this region and so their fitnesses by picking new products forme their listing.

3. Those with incomes that strikingly surpass the average income are considered to be overextended with respect to the liquidity of the market and consequently, tend to reduce their incomes.

3.1.4 Vendors' dynamics of the constrained market activity

The same analysis can be applied to the constrained markets. In this case, the reduction in the listing diversification increases the correlation between high fitness and high incomes, as it can be seen in Figure 3.10.

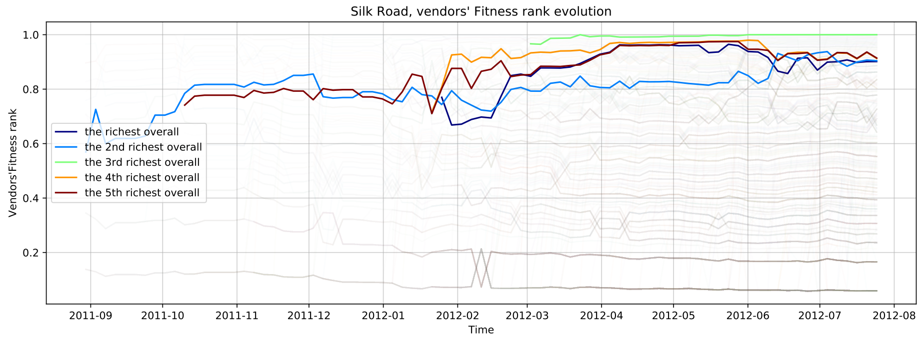
Figure 3.12 shows how the arrow plot and the streamlines of the market dynamics change in case we concentrate the analysis on the constrained market activity. In this case, the binary matrix is almost *flat belly*, as shown in Figure 3.4, and the fitness seems to better capture the competitiveness of the vendors. The first proof of that is given in Figure 3.10 and in Figure 3.11: we can appreciate how the richest vendors of the overall market activity are actually those with a high fitness rank, and ulteriorly, their evolution to an increasing income follows a coherent increment of their fitnesses. It is worthy to notice that the constrained dataset has exactly the same richest vendors that were the richest over the total dataset. Moreover, we can appreciate how a reducing the dataset to the top twenty products can redistribute the fitnesses of the vendors, due to a different categorization of their degree of diversification. As it regards the coarse-grained vendors' dynamics, Figure 3.12 shows how the vendors with the highest fitness and a low income are pushed to increase their revenues more than in the previous case of the total market activity, Figure 3.9. It seems that the basic definition of economic competitiveness, as the capability of diversifying and producing the most complex products, results completely less intuitive for general and dynamical markets, while it works better for more settled ones, as the global economy or the core activity of the Silkroad Darknet market (97.8 % GDP).

3.1.5 Interpretations of the Silkroad'results

In the contest of the global economy, if a country is suddenly able to export a rare product, such as spaceships, the algorithm will consider that country able to produce almost everything else, and consequently, one of the fittest. In the global economy, this is actually true. If a country can compete in a market sector it will do it, reducing the possibility of oligopolies and moving the belly of the binary matrix to be outward. On the contrary, in the Darknet market, if a vendor is capable of selling the rarest product, the algorithm considers him to be able to produce also the less complex ones, but in reality that doesn't mean that he will actually cover all the other market sectors. Looking back at the ordered binary matrix, Figure 3.1, we can appreciate how just a few of them really diversify their listing increasing their fitness, but this behaviour it is not reflected by an increase in their revenues, as supported by Figure 3.7. In the Darknet market diversification doesn't have an immediate fruitful impact on the market, differently from the global economy. We can try to give an explanation looking at the two different markets' demands; in the global economy, countries end up exporting products only if there is a sustained global demand, whereas, in Darknet markets most of the listed products still have to face a wave of customer adoption, hence they don't have an impact on vendors' revenues. Indeed, as noticed in 1.6, sectors that mostly move the market's revenues are only a few.

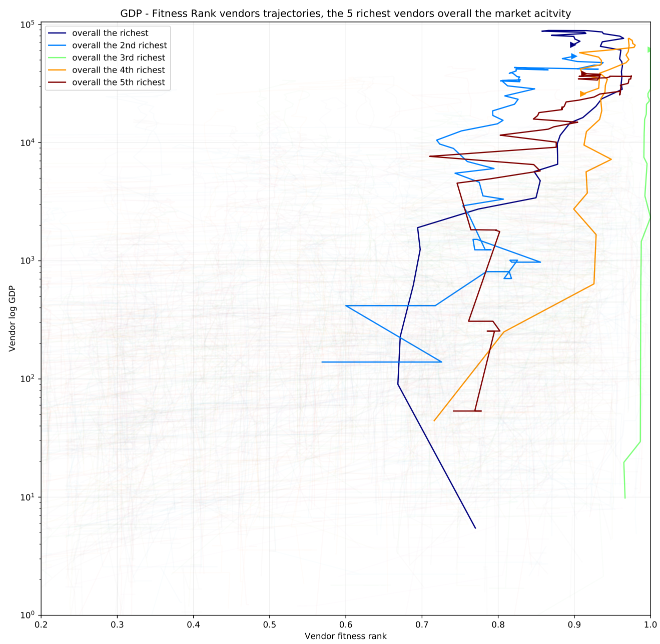
How can we correlate the fact that incomes of vendors are all strictly related to few products with the fact that in the mean field trajectories, all the vendors are able to earn almost the same amount of money, despite their fitness?

Figure 3.10: Silk Road (97.8 % GDP), evolution of the vendor's fitness ranking



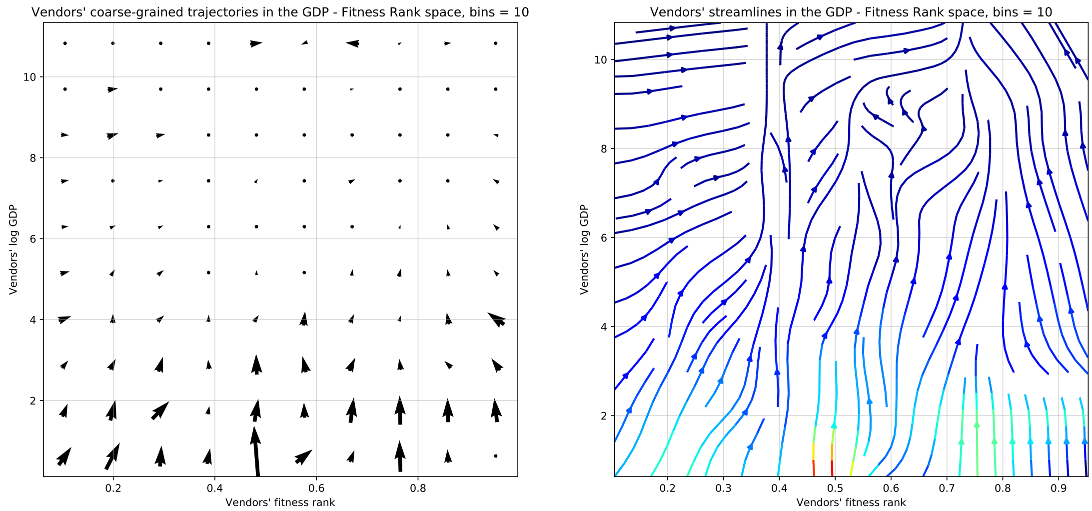
Evolution of the ranking of the vendors' fitnesses in the Silk Road Darknet market. The dataset is related to the constrained market activity of Silk Road (97.8 % GDP). Of all the 849 vendors we have highlighted the five richest vendors over the total timeline of the market activity. The entrance of new vendors is clearly visible in the figure.

Figure 3.11: Silk Road (97.8 % GDP), vendors' trajectories in the GDP - Fitness Rank space



Vendors' trajectories in the GDP - Fitness Rank space. The dataset is related to the constrained market activity of Silk Road (97.8 % GDP). Of all the 849 vendors we have highlighted the five richest vendor over the total timeline of the market activity.

Figure 3.12: Silk Road (97.8 % GDP), the arrow-plot and the streamlines of the vendors' trajectories in the GDP-Fitness space



The plot on the left is the arrows plot of the coarse-grained version of the vendors' trajectories in the GDP-Fitness space. While the plot on the right represents the streamlines of the vendors' trajectories in the GDP-Fitness space. The size of the arrows and colours state the norm of the incremental vectors in each bin, with an increasing value going from blue to red. We have considered only the boxes that contained more than 3 points. The dataset is related to the constrained market activity of Silk Road (97.8 % GDP).

How can we correlate the fact that incomes of vendors are all strictly related to few products with the fact that in the mean field trajectories, all the vendors are able to earn almost the same amount of money, despite their fitness?

An interpretation can be that due to competition, there is indeed an equilibrium characterised by an optimal fitness and sales, the basin of attraction shown in Figure 3.9 and Figure 3.12. It can be that the analogy with countries should not necessarily apply in the case of Darknet marketplace, as the space of solutions and the potential for creativity and growth are very different from those of countries in the global economy. In fact, in the Darknet market case, the only products that move significantly the market are the ubiquitous ones and almost all vendors are able to sell them. And, secondly, we should remind that the common average rating between the vendors and the large trustworthiness of customers in the escrow system, make the customers willing to buy those few products from whoever is able to sell them, increasing the revenues of all the vendors, despite their fitnesses.

The fact that the market is concentrated on a few numbers of products exaggerates the ubiquity of some classes, and give rise to two other interpretations of the results.

- The first one is that the market is truly immature and has been through just to the first wave of market adoption, strictly related to those mainstream products.
- A more subtle interpretation can be that this is not a real delay in the demand but just a wrong categorization of the latter, probably due to a wrong categorization

of the products in the dataset. There is a clear variety in the degree of specificity of the products categories, we argue that this sloppy categorization could introduce misleading results.

We can clearly see that the most important products are usually those categorized in a general way, this evident error could be the reason why the incomes are so disproportionally distributed over the products category.

We aren't able to distinguish the typology of the most appreciated products, such as *Weed*, *Books* or *Pills*, while on the other hand, we are able to distinguish, with an extreme degree of specificity, some psychedelic substances or some variety of edible magic mushrooms, that inevitably makes them rarer and consequentially fitter. It is like if we are in a mall and we try to make a comparison between the ubiquity of the category *Shoes* and the ubiquity of the category *Stylographic pen, with navy ink*.

In this interpretation, the over dominance of some classes of products is just due to their sloppy specificity. This can be one of the reasons for the unexpected vendors' dynamics in the total dataset and the better ones for the constrained case.

3.2 Results on Agora

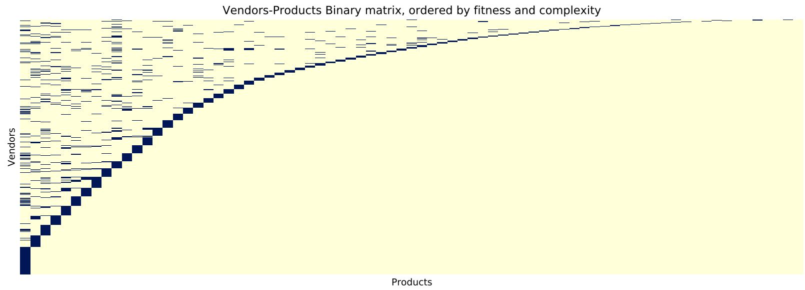
In order to understand the generality of the features detected in the Silkroad Darknet market, we inspected the Agora Darknet market in the same way. The Agora Darknet market offers the possibility of analysing a network that doesn't suffer the evolutional characteristics of Silk Road. The number of vendors and products is almost stable in each time window, between 700 to 900 vendors, and 63 to 77 products. Moreover, the Agora dataset doesn't suffer the sloppy categorization of the products. The number of products is in fact reduce, and they all share the same degree of specificity.

Unexpectedly, the results are really similar and they let us think that the algorithm has underlined a general feature of the anonymous interactions between agents in the Darknet marketplaces.

The binary matrix, shown in Figure 3.13, is extremely inward belly and it doesn't permit to extrapolate any crossing country point. Consequently, the convergence behaviour is exponential to zero values for all the vendors except for one of them, Figure 3.14, Figure 3.15.

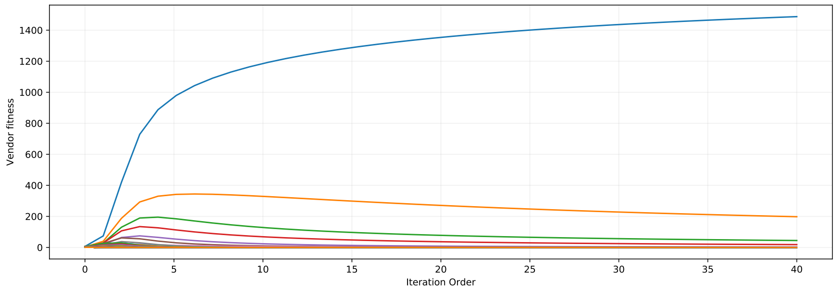
The evolution of the vendors' fitness rank, given in Figure 3.16, shows how, like in the general Silk Road case, the vendors that have performed better in terms of revenues are not those that have had the best fitness rank. Moreover, the arrow plot and the streamlines, in Figure 3.17, are also quite similar. The lateral regimes, related to the convergence into the basin of attraction, are even more defined for the Agora Darknet market, probably due to a more stable market and to a larger number of data points, aka transactions. It is evident how, even in this more stable, the push-down regime is even stronger, pointing out how difficult is for the vendors to keep increasing their revenues over the average value and how ineffective is to maintain an over diversification in order to be competitive. As a final remark, we are moved to consider that the inefficacy of the algorithm is neither related to the evolutional characteristics of the market, such as the variation in the nodes number, nor to its bad categorization of the products.

Figure 3.13: Agora, the ordered binary matrix



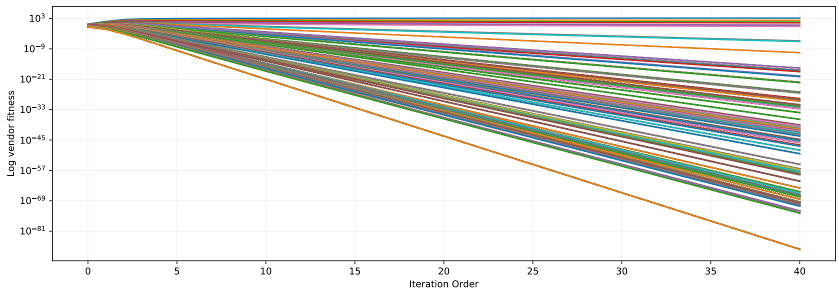
The ordered binary matrix of the Agora Darknet market. The dataset is related to the total market activity.

Figure 3.14: Agora, convergence behaviour of the vendors' fitnesses



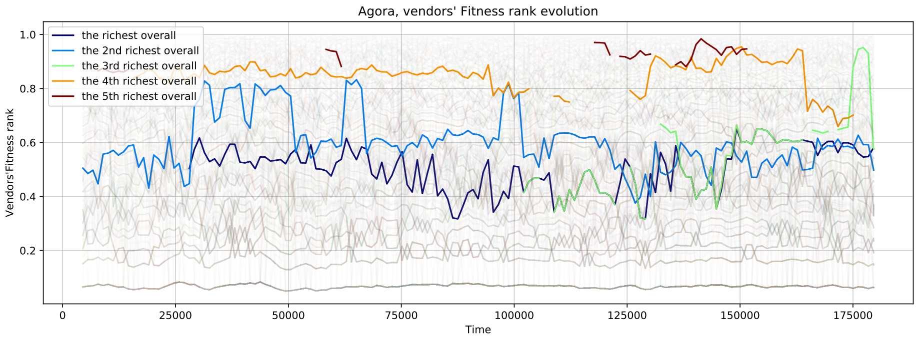
Convergence behaviour of the vendors' fitnesses. The dataset is related to the total market activity of the Agora Darknet market.

Figure 3.15: Agora, logarithmic convergence behaviour of the vendors' fitnesses



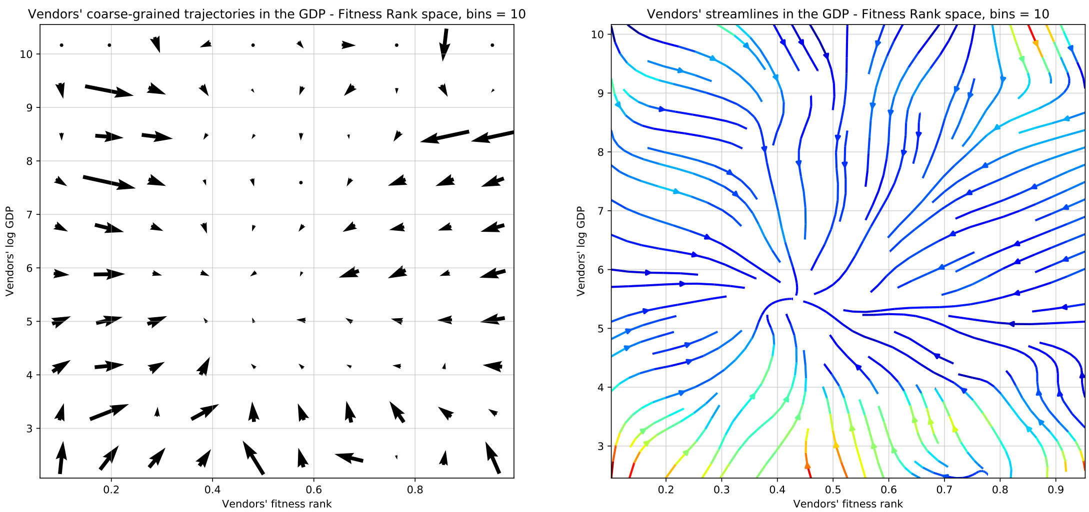
Logarithmic Convergence behaviour of the vendors' fitnesses. The dataset is related to the total market activity of the Agora Darknet market.

Figure 3.16: Agora, evolution of the vendor's fitness ranking



Evolution of the ranking of the vendors' fitnesses in the Agora Darknet market. Of the 1752 vendors we have highlighted the five richest vendors over the total timeline of the market activity. The entrance of new vendors is clearly visible in the figure.

Figure 3.17: Agora, the arrow-plot and the streamlines of the vendors' trajectories in the GDP-Fitness space



The plot on the left is the arrows plot of the coarse-grained version of the vendors' trajectories in the GDP-Fitness space. While the plot on the right represents the streamlines of the vendors' trajectories in the GDP-Fitness space. The size of the arrows and colours state the norm of the incremental vectors in each bin, with an increasing value going from blue to red. We have considered only the boxes that contained more than 3 points. The dataset is related to the total market activity of the Agora Darknet market.

Chapter 4

Tripartite network's results

4.1 Case 1, Customers treated as singular agents

In the previous Chapter 3, we have seen how Pietronero et al's metrics lead to controversial results, probably due to the lack of information as it regards the role of customers' demand. On the following, we are going to discuss the results related to the first generalisation that we have proposed in Section 2.4.1. We repropose here the set of equations and we remind that the main variation is that we added the fitness of the customers, F_c.

$$\begin{cases} \tilde{F}_p^{(n)} = \left(\dfrac{1}{\sum_{v=1}^{N_v} M_{vp} \dfrac{1}{F_v^{(n-1)}}} \right) & (4.1) \\[2em] \tilde{F}_v^{(n)} = \displaystyle\sum_{c=1}^{N_c} M_{vc} F_c^{(n-1)} & (4.2) \\[2em] \tilde{F}_c^{(n)} = \displaystyle\sum_{p=1}^{N_p} M_{cp} F_p^{(n-1)} & (4.3) \end{cases}$$

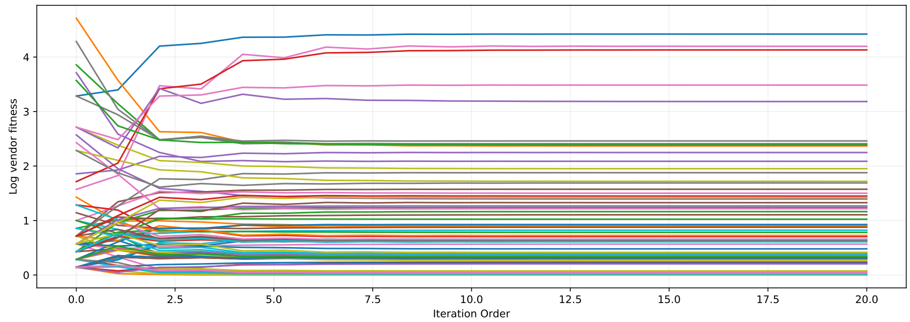
4.1.1 Convergence and reintroduction of the cardinality

The implementation of the generalised Fitness-Complexity Algorithm still makes use of a rolling window in order to smooth and accomplish for the lack of data. Moreover, as previously stated we have decided to consider only customers that have spent more than 10 BTC and the top twenty products. This reduction of the Silk Road dataset still permits us to consider the 85 % of the vendors, the 55 % of the customers and the 97.6 % of the market GDP.

The first result concerns the new convergence of the vendors' fitness to values different from zero. This first achievement permits us to retain the cardinality of the fitnesses, a property that we have always lost trying to straightly apply the original Fitness-Complexity Algorithm to a bipartite network of vendors and products.

The property of convergence to non zero values has been tested numerically. In Figure 4.1 we can appreciate an example of convergence.

Figure 4.1: Tripartite generalisation of the metrics (Case 1). Constrained Silk Road (97.6 % GDP), logarithmic convergence behaviour of the vendors' fitnesses



Logarithmic convergence behaviour of the vendors' fitnesses, using the tripartite generalisation of the metrics. The dataset is related to a constrained market activity of Silk Road (97.6 % GDP). It takes into account the top 20 proficuous products that constitute the 97.6 % of the market revenues.

4.1.2 Vendors' dynamics of the constrained market activity

Introducing the customer as a singular agent we increase the correlation between fitness and incomes. In this case, the fitness better captures the competitiveness of the vendors.

The coarse-grained vendors' dynamics utilizing this new set of equations, we have finally trajectories that take into account the cardinality between the values of fitness, exactly like in the case of the bipartite network of countries and products. This new feature change completely the phase space, which in this case has been plotted using log-scaling both the fitnesses and the incomes. Figure 4.2 underlines a higher correlation between the fitnesses and the incomes with respect to that obtained with the original Fitness-Complexity Algorithm.

We can appreciate mainly three different regimes:

1. The diagonal regime: it underlines the optimal strategy to increase the revenues while optimally reducing the diversification of products in the market.

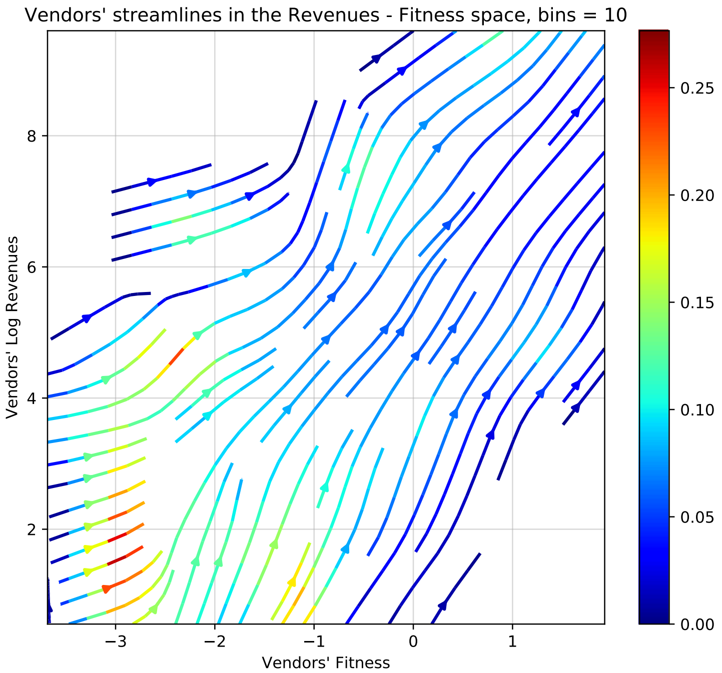
2. The left regime with respect to the diagonal: Vendors with a lower fitness tend to first adjust their fitness and successively increase their revenues.

3. The right regime with respect to the diagonal: Vendors with a higher fitness tend to first adjust their fitness and successively increase their revenues.

We conjecture that the region with a larger magnitude in the incremental vectors (bigger arrows and reddish colour) suffer from an artefact. It seems that some vendors are more susceptible than others to the customers' demand and present some large process of mean reversion. We think that these processes alter the arrow plot, covering almost completely the variations in the y-direction.

Unfortunately, the dataset related to the Agora Darknet market doesn't contain hash-encrypted information of the customers, thus it doesn't permit us to generalize the features obtained on the tripartite network of Silk Road to some general feature of the Darknet marketplaces.

Figure 4.2: Tripartite generalisation of the metrics (Case 1). Silk Road (97.6 % GDP), the streamlines of the vendors' trajectories in the GDP-Fitness space



The plot represents the streamlines of the vendors' trajectories in the GDP-Fitness space. The size of the arrows and colours state the norm of the incremental vectors in each bin, with an increasing value going from blue to red. We have considered only the boxes that contained more than 5 points. The dataset is related to the constrained market activity of Silk Road (97.6 % GDP)

4.2 Case 2, customers treated as an informative noise feedback

Hereafter, we are going to discuss the results related to the second generalisation that we have proposed in Section 2.4.2. We repropose here the set of equations and we remind that the main variation is that we have kept the framework of the Fitness-Complexity

Algorithm, with the introduction of the feedback rates that customers need to release after each transaction.

$$\begin{cases} \tilde{F}_p^{(n)} = \left(\dfrac{1}{\sum_{v=1}^{N_v} M_{vp} \dfrac{1}{F_v^{(n-1)}}} \right) & (4.4) \\ \tilde{F}_v^{(n)} = \displaystyle\sum_{p=1}^{N_p} M_{vp} F_p^{(n-1)} \tilde{R}_{vp}^{(t)} & (4.5) \\ \tilde{R}_v^{(t)} = \displaystyle\sum_{p=1}^{N_p} R_{vp}^{(t-1)} + \frac{1}{t} (R_{vp}^{(t)} - \tilde{R}_{vp}^{(t-1)}) & (4.6) \end{cases}$$

We have thus developed the same analysis previously explained.

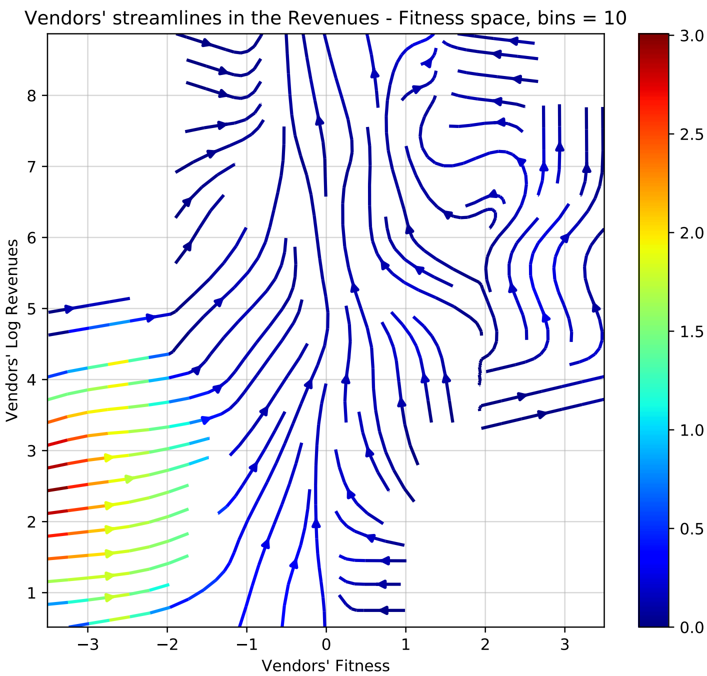
This new generalization doesn't cure completely the problem of an exponential decay, even if reduces the number of exponential curves. We have further proceeded retaining fitness's values that don't converge to zero. Reducing the number of datapoints while maintaining both cardinality and ordinality of such values.

Figure 4.3 shows an improvement of the metrics informativeness with respect to the fully bipartite Fitness-Complexity Algorithm, the results of which are shown in the previous Chapter 3. The left side shows how vendors need to increase their fitness first in order to increase their revenues. In this new framework, it means that vendors should increase their diversification over products and obviously head a toward an increase of their average feedback rate. However, the right side of the coarse-grained trajectory space shows a more chaotic situation. It seems that the vendor's fitness fails to describe their level of competitiveness.

Now the interpretation can be that, formally, the network that we are considering is still bipartite, and the results are only slightly modulated by the introduction of the feedback rates. However, keep considering diversification over products as the major factor of competitiveness is still wrong, an evidence of which is given by the failure of every metrics that make use of it (Fitness-Complexity Algorithm 2.2.2 and the last version modulated by feedback rates 2.4.2.)

In the next Chapter 5 we will further analyze the different impact of diversification over customers and products, understanding even more the striking importance of the former strategy.

Figure 4.3: Tripartite generalisation of the metrics (Case 2). Silk Road (97.6 % GDP), the streamlines of the vendors’ trajectories in the GDP-Fitness space



The plot represents the streamlines of the vendors’ trajectories in the GDP-Fitness space. The size of the arrows and colours state the norm of the incremental vectors in each bin, with an increasing value going from blue to red. We have considered only the boxes that contained more than 5 points. The dataset is related to the constrained market activity of Silk Road (97.6 % GDP).

Chapter 5

Theoretical framework II

5.1 Production capabilities and Trade capabilities

Following the lead of the "theory of capabilities" in the economic complexity literature, [10, 9], we introduced a model of economic growth moving from the tripartite network of countries-capabilities-products to the tripartite network of vendors-capabilities-products, subsequentially applied to the case study of Darknets markets. As well as exported products are considered to discount all the information on countries' capabilities, we started assuming that products discount all the information on the competitive capabilities of vendors. Originally, such an assumption was based on a first essential result of the analysis, the degree of diversification over the set of products is intrinsically related to the competitiveness of a country.

However, in our economic networks, we have not observed such a behaviour. The most successful vendors don't need to diversify over a large set of products in order to increase their gains, namely products are not an immediate reflection of a competitive capability. On the contrary, we have observed that the feature that correlates with the highest gainer is not diversification over products, instead, it is diversification over customers. Indeed, we have observed that the most successful vendors only choose a small set of appetible products and then spread their sells over a wide set of customers. The success of the generalized algorithm let us conjecture that both customers and products discount all the information on capabilities, and leads us to following statements:

- Capabilities may be of two kinds: Production capabilities (related to the production structure) and Trade capabilities (related to the selling structure)
- Vendors tend to build on existing capabilities to move into new productive activities and to attract new customers.
- This induces a path dependent diversification process.

Having noticed that, how is it possible to model such a process?

Hereafter we introduce and readapt techniques that come from the economic complexity literature, with the purpose of backtesting the previous results given by the generalized

Fitness-Complexity Algorithm, and, furthermore, justify the superiority of a focused business strategy (4-5 products and lots of different customers) with respect to a dynamical strategy that updates products and increase diversification. A result that disproves the merit of a product diversification strategy for companies/organization in favour of a Ricardo's principle of specialisation, in line with what observed by Pietronero et al. in [18].

Methodology speaking, we take again inspiration from the work of Hidalgo & Haussmann, [10, 9], and, starting with the concept of Product Space, we have been able to build the analouge Client Space of our market, as the network of relatedness between customers.

In Chapter 6, we clearly appreciate the results of such backtest. Indeed, the analysis of vendors' diffusion on both Spaces shows that:

- the most successful vendors largely populate the Client Space and strategically (maybe unconsciously) navigate such a space. More in details, as soon as the Client Space reach a coherent and stable topology, top vendors start to develop patterns of diffusion with centralities, defined in 5.2.2, that correlates with their level of revenues. Namely, top sellers tend to increase their centralities while increasing their revenues and viceversa. We conjecture that a successful degree of Trade capabilities is responsible for a successful high centrality diversification over customers.

- the most successful vendors don't need to update their baskets of products in order to impose their competitive power, hence they tend to show a quasi-static degree of diversification over the Product Space. Coherently, we conjecture that, in our economic network, a successful degree of Production capabilities is responsible for a quasi-static diversification over products, that may cover an hidden process of specialisation on that set of products (new models, better technologies..)

In this new framework, the most successful vendors tend to diffuse through the Client Space reaching only new customers, that are close to customers that they already have (nearest neighbour connection). The same for products in the Product Space. Such results are an evidence of this two different set of capabilities; again, Trade capabilities and Production capabilities.

An example of such a pattern of diffusion comes from the literature and may be useful in order to conceptualize. The Product Space can be visualized with the following metaphor: each product can be considered as a tree, and the collection of all products to be a forest. A country or a vendor in our case can be considered as a monkey that populates the forest. A vendor populates the Product Space depending on its products' basket and, in the same manner, populates the Client Space depending on its clientele. For such "economic monkeys", the process of growth means moving from a poorer part of the forest, where the trees bear little fruit, to a better part of the forest. To do this, the monkeys must jump distances; that is, in the case of a vendor in the Product Space, redeploy (physical, human, and institutional) capital to make new products or, in the case of a vendor in the Client Space, reinvest (marketing, products..) to target specific clients. Traditional economic theory disregards the structure of the forest, assuming that there is always a tree within reach. However, if the forest is not homogeneous, there will be areas of dense tree growth in which the monkeys must exert little effort to reach new trees and sparse regions in which jumping to a new tree is very difficult. In fact, if some areas are very deserted, monkeys may be unable to move through the forest at all. Therefore, the structure of the forest and a monkey's location within it dictates the monkey's capacity for growth; in terms of economy, the topology of the Product Space and the Client Space impact the vendors' ability to begin selling new goods or reaching new customers.

5.2 Methods

For the sake of conciseness, we refer to previous reports for the Fitness-Complexity Algorithm. Hereafter, we only introduce the Space metrics.

5.2.1 The Product Space and the Client Space

In the spirit of [9], we introduce here a straightforward transmutation of the so-called Proximity matrix, and its Product Space visualization, on the case study of Darknet marketplaces. Such a metric calculates the minimum probability that a vendor sells the product p given that it sells the product p' and vice-versa. Thus, in our analogy, the proximity matrix of the vendors-products network calculates the relatedness between products. Two products are defined to be proximate if they are both sold from the same vendors. For the sake of notation, we remember that M_{vp} is the usual ordered binary matrix of vendors and products, which has been explained in Section 2.1.1.

$$Proximity = \phi_{pp'} = \frac{\sum_v{}^{N_v} M_{vp} M_{vp'}}{max(\sum_v{}^{N_v} M_{vp} \ , \ \sum_v{}^{N_v} M_{vp'})} \tag{5.1}$$

The Proximity matrix can be further visualized as a network, the so-called Product Space; a weighted network in which nodes are the different products and the weighted edges are probabilities of co-trade, a sort of adjacency matrix. In this context, the Product Space is used to analyze how each vendor updates and adjust its range of products over time.

In order to understand how the customers'demand shapes the succes of vendors, we conjecture an analogue of the Product Space, built on the Proximity matrix given by the vendors-customers network; the Client Space. (Equation 5.1 with a $\phi_{cc'}$ based on the vendors-customers binary matrix M_{vc}). Again, the idea is that two clients are proximate in the network if they both buy from the same vendors. It follows that the periphery of the Client Space represents customers that buy less and only from few vendors, while, on the other hand, the core of the Client Space represents the cluster of customers that buy more and from lots of different vendors. We recall that the defintion of periphery and core of the network are supscetible of the centrality measure used to define them.

Figure 5.1 shows the Product Space of the entire Silk Road's dataset. It represents the maximum spanning tree of the "Force Atlas" network layout of the Product Proximity matrix. The network has been extracted through Gephi and the different colours have been computed with a community detection algorithm [19], with a resolution parameter of 1.5. On the other hand, Figure 5.2 shows the Client Space of a constrained Silk Road's dataset

(2900 most spendthrifts). It represents the maximum spanning tree of the "proportional Yifan Hu" network layout of the Client Proximity matrix. Again, the network has been extracted through Gephi and the different colours have been computed with the same community detection algorithm [19], with a resolution parameter of 10. Through this network visualization, it is possible to appreciate different clients communities/clusters susceptible to buying from the same vendors.

In the next Chapter 6, we will refer to a smaller subset of clients, in order to keep the computations treatable.

5.2.2 Vendors' centrality and revenues

Starting from this arguments, a valuable tool to indagate the different features of vendors diffusion is given by the notion of centrality. Indeed, the centrality of a single node gives us an idea of its importance in the Client Space. Accordingly, the average centrality of a group of nodes, the ones populated by the vendor, gives us an idea of the importance of a vendor, i.e how much central he is in the Client Space.

Notation

$k_v(t)$:= vendor's centrality at time t $k_c(t)$:= client's centrality at time t

$$k_v(t) = \frac{\sum_{c_j} k_{c_j}(t)}{N_{c_j}}, c_j \; client \; of \; v \tag{5.2}$$

It is now legitim to question if the centrality of a vendor shapes his revenues, or, with less audacity if revenues correlate with the vendor's centrality. In this respect, we first developed an analysis that includes different centrality measures (betweenness centrality, closeness centrality and Katz centrality), the definition of which has been given in Appendix A 6.3. Furthermore, for all vendors in each time window, we have indagated the correlation between the average vendors' centralities and the vendors'revenues. Results are shown in Chapter 6, Figure 6.3.

Moreover, in Chapter 6, Section 6.3 we propose an analysis that gives us a feeling of how much being central in the Client Space increases revenues than just having a large number of clients with a random position. Namely, we study the statistical performance of multiple linear regressions of all the models that we consider to be relevant; a mixture of the four possible dependent variables (the three vendors' centrality measures and the vendors' number of clients) and the unique independent variable (vendors' revenues). Indeed, a higher statistical significance of models that includes centrality variables would underline that positions of customers in the Client Space aren't equally profitable. Such an evidence would legitimize the introduction of the Client Space as a methodological tool of analysis.

Figure 5.1: Silk Road, the Product Space

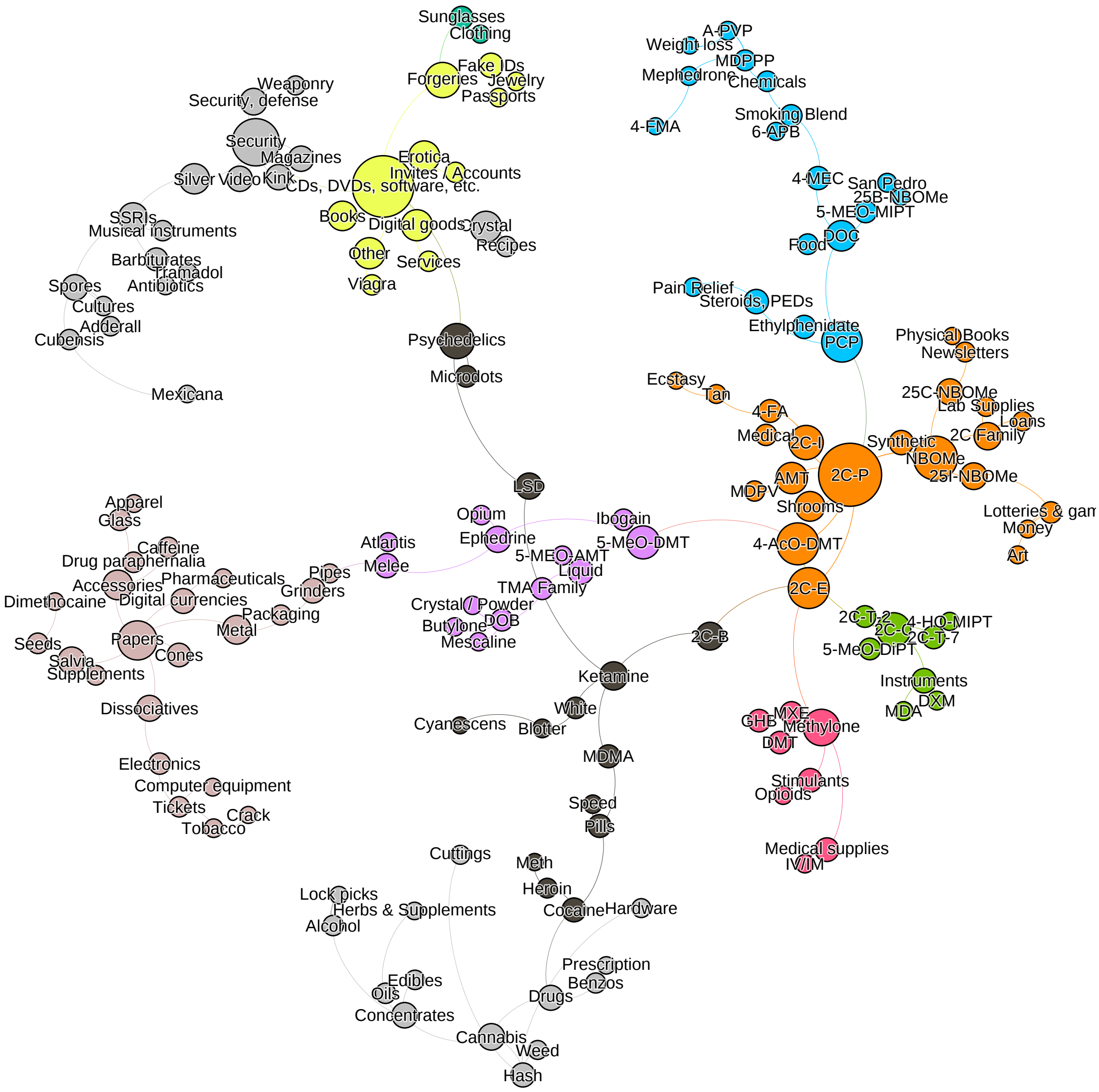
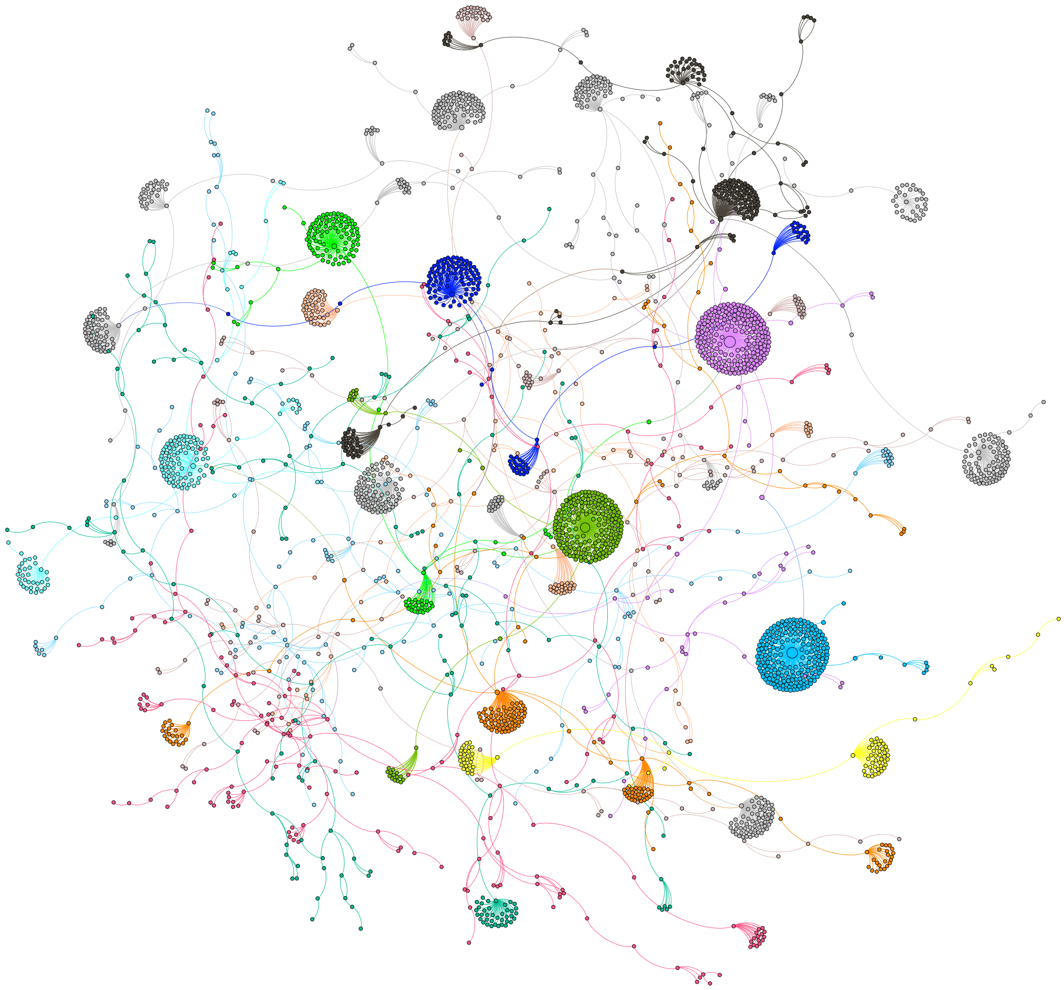


Figure 5.2: Silk Road, the Client Space (the 2900 most spendthrifts clients)



Chapter 6

Product Space's and Client Space's results

6.1 Dataset

On the following, we have decided to concentrate the analysis over a constrained version of the dataset, reducing it to:

- the thirty most relevant products by total cash flow, over the market timeline.
- all the customers that have spent more than 300 BTC. In order to reduce the number of different aliases of customers and to concentrate on the most relevant set of customers (424).
- all the vendors that have traded the selected products and customers (669).

On the long run, we have detected that some vendors have constantly beaten the market, while others have had a disruptive entrance, and from there have slowly climbed the rank of the most profitable. Hereafter, we will focus on the set of most profitable vendors in order to detect patterns of the best business strategies.

6.2 Product Space and Client Space: different patterns of diffusion

The procedure described in section 5.2.1 permit us to analyse the different pattern of diffusion of top performer vendors on the two spaces (Product Space and Client Space).

From the second column of Figure 6.1, it is evident how the most successful vendors don't need to update their baskets in order to incentivize their activities, which flourish even with a small set of desirable products (3 to 5 mainstream products). Such a result seems to underline that products are not the expression of a set of competitive capabilities; it is not true that the more a vendor is diversified the more is competitive. Moreover, this argument is in line with the failure of the application of the Fitness-Complexity Algorithm on the bipartite network of vendors and products, extensively discussed in previously.

It follows, coherently, that the only way to get reacher is to sell more. This can be achieved in two ways: by selling more to the same set of clients or by enlarging such a set. The evidence of the latter is given by the first column of Figure 6.1. The most successful vendors have a spread diffusion over a large set of clients. Furthermore, two different features may be detected:

- As the monkey of the metaphor (Section 5.1), vendors tend to *jump* between nearby (connected) customers. Which imply, that such customers share some sort of *demandcapabilities* (which may consist of common interests, susceptibility to a common marketing strategy and so on..)
- The most successful vendors are more central, i.e tend to populate the core of the Client Space. Such a feature, extensively tested with the correlation procedure of Section 5.2.2, the results of which are shown in Figure 6.3, let us conjecture that such a dynamic may be an indicator of a successful business strategy. Trading with core-customers seems more profitable in the long run.

6.2.1 Video visualization of the Top Seller diffusion

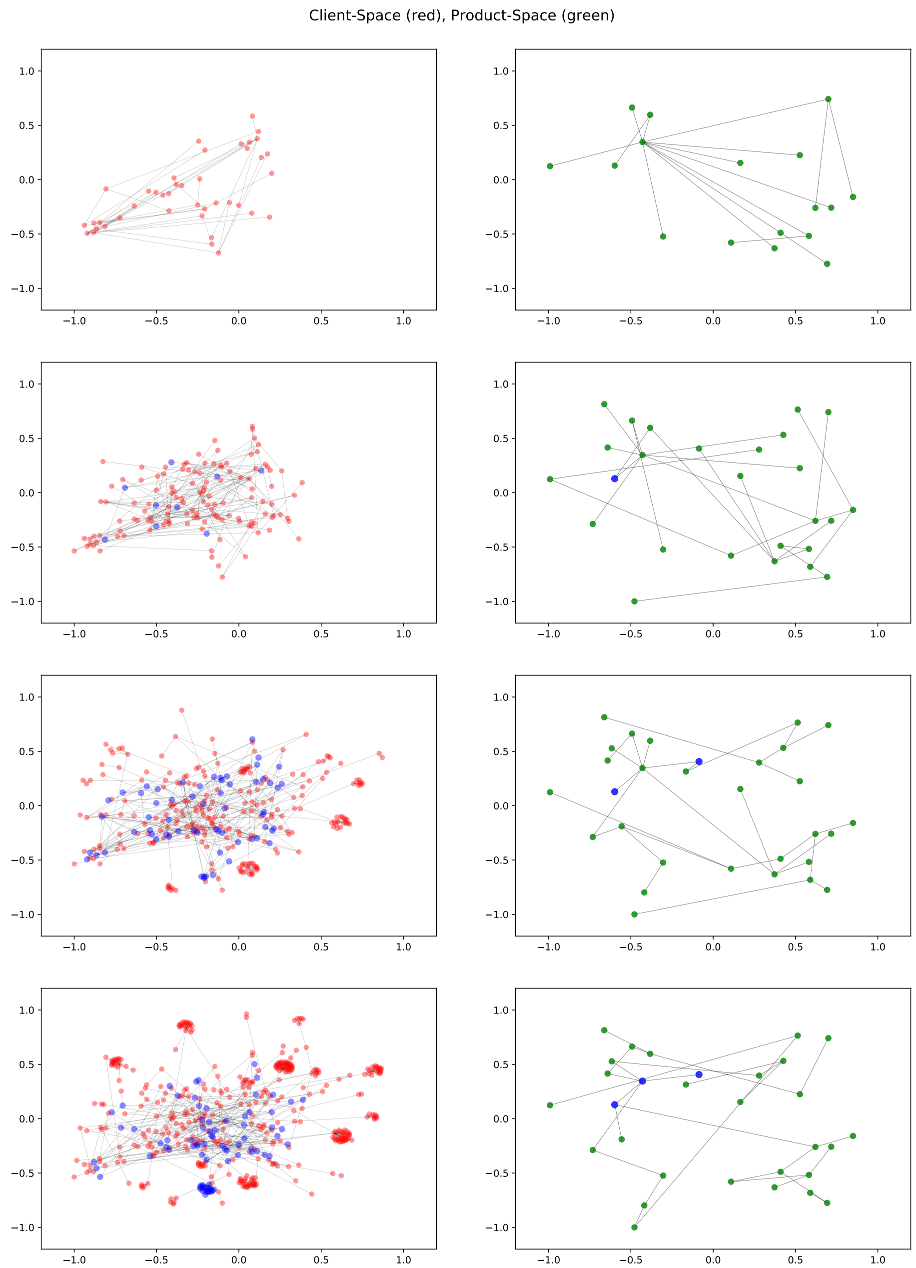
To visualize such a video 6.2, is required to open the PDF file through Adobe Reader.

Hereafter, a video of the Top Seller diffusion over the Client Space and the Product Space. As previously explained, we can visualize how top sellers tend to express:

- a high centrality diversification over customers
- a quasi-static diversification over products

The right side of the video represents the diffusion of the top seller on the Client Space. The topology of the Client Space varies along the market activity and reach a stable structure only after the first half of it. Red nodes, that appear and disappear, are general clients, while blue nodes are the clients that, in that frame of the market activity, have consistently ($RCA > 1$) traded with the Top Seller. We can appreciate how blue nodes tend to be concentrated and how the *entering* blue nodes are usually connected with previous blue nodes. We conjecture that this is an evidence of the existence of the aforementioned Trade capabilities. Sometimes it appears that there is not a straight connection between the appearance of a *entering* blue node and a cluster of already existent blue nodes, however, we need to remember, that we are dealing with the maximum spanning tree of the network layout, which only shows some edges, the ones that maximize the edges' weights of the spanning tree.
As it regards the Product Space, the left side of the video, green nodes are general products and, again, blue nodes are the products that has been consistently ($RCA > 1$) traded by the Top Seller in that frame of the market activity. It is clear how blue nodes are strictly connected and thus similar in term of Production capabilities. The quasi-static diversification over products is enhanced by such visualization.

Figure 6.1: Silk Road, Client Space and Product Space, example of diffusion of the 2nd top seller by revenues ("2ec42a461e")



The first column (red and blue nodes) represents its evolution on the Client Space, among four different time windows. On the other hand, the second column (green and blue nodes), represents its evolution on the Product Space, among the same four different time windows. In both of the series of networks, the blue nodes are the clients (or products) that the seller "2ec42a461e" can reach (or trade) with a comparative advantage ($RCA > 1$), the red and the green ones vice-versa. The entrance of new customers and new products justify the different topologies among time windows.

Figure 6.2: Silk Road, the Top Seller diffusion, comparison between Client Space's diffusion and Product Space's diffusion, (video)

(Loading Video...)

6.3 Multiple Linear Regressions: The best Centrality Model

We need to notice that the other road to success, maintaining a small set of customers while increasing revenues, is limited by the size and frequency of trades that are naturally constrained by the illicit nature of the market and the willingness to not arouse suspicions. Consequently, the only viable option is to enlarge the set of clients.

In this respect, we propose a set of different analysis that studies how revenues correlate with the number of vendors and with different notions of centrality. Following results expressed in Figure 6.3, we have developed a set of Multiple Linear Regressions on a unique time window, corresponding to the second, and more relevant, half of the market activity. In order to understand if clients centralities have a statistical impact on the vendor profitability we have compared different nested models, that gradually include this information, hereafter treated as model's parameters.

One model (Model 6.1) includes all the centralities plus the number of clients, three models (Model 6.2, Model 6.3 and Model 6.4) with a single centrality plus the number of clients, and, a last model (Model 6.5) that only includes the number of clients without taking care about their position in the Client Space.

Notation :
x_1 : vendor's betweenness centrality,
x_2 : vendor's closeness centrality,
x_3 : vendor's Katz centrality,
x_4 : vendor's number of clients,
y : vendor's revenues.

Models :

$$1) : y = a * x_1 + b * x_2 + c * x_3 + d * x_4 \tag{6.1}$$

$$2) : y = a * x_1 + d * x_4 \tag{6.2}$$

$$3) : y = b * x_2 + d * x_4 \tag{6.3}$$

$$4) : y = c * x_3 + d * x_4 \tag{6.4}$$

$$5) : y = d * x_4 \tag{6.5}$$

Multiple Linear Regressions on the previous models give the results extensively reported in Appendix B 6.6.

The highlights are:

- Model 6.1 has the highest *Adj.R squared* (0.538), with each $p - value$, one for every variable, way smaller than the standard significance level of 0.05. (More in details, $p - value(x_1) = 2.715e - 04$, $p - value(x_2) = 5.131e - 05$, $p - value(x_3) = 5.236e - 12$, $p - value(x_4) = 4.480e - 78$).

- Model 6.5 has the smallest *Adj.R squared* (0.493) and the smallest $p - value$. ($p - value(x_4) = 9.567e - 94$)

All the other models, 6.2, 6.3, 6.4, have an *Adj.R squared* between the previous two values ($0.492 < Adj.\ Rsquared < 0.537$), and $p - values$ significantly small, except for Model 6.2. Indeed, in such a case, the $p-value$ of the x_1 variable cannot be accepted, as it is bigger than the standard significance level of 0.05 ($p - value(x_1) = 0.101$). Therefore, Model 6.2, the one that include betweenness centrality plus the number of clients, cannot be accepted.

Until now, what these results mean?

- The number of clients is the most significant dependent variable, as it has always the smallest $p - value$.

- However, closeness centrality and Katz centrality increase the predictivity of the model. In other words, model that includes this information, 6.3 and 6.4, are more precise than Model 6.5.

Let's now understand which models, between 6.1, 6.3 and 6.4, has the highest statistical significance with respect to the Model 6.5.

We have already seen how Model 6.1 has the highest *Adj.R squared*. While, between the 6.3 and 6.4, the one with higher *Adj.R squared* (0.509) is Model 6.4, the Katz centrality one.

However in order to understand how better than Model 6.5 they perform, we followed Wilk's theorem for nested models [22], and we implemented the $Log\ Likelihood\ Ratio\ Test$ [21].

Results shows that:

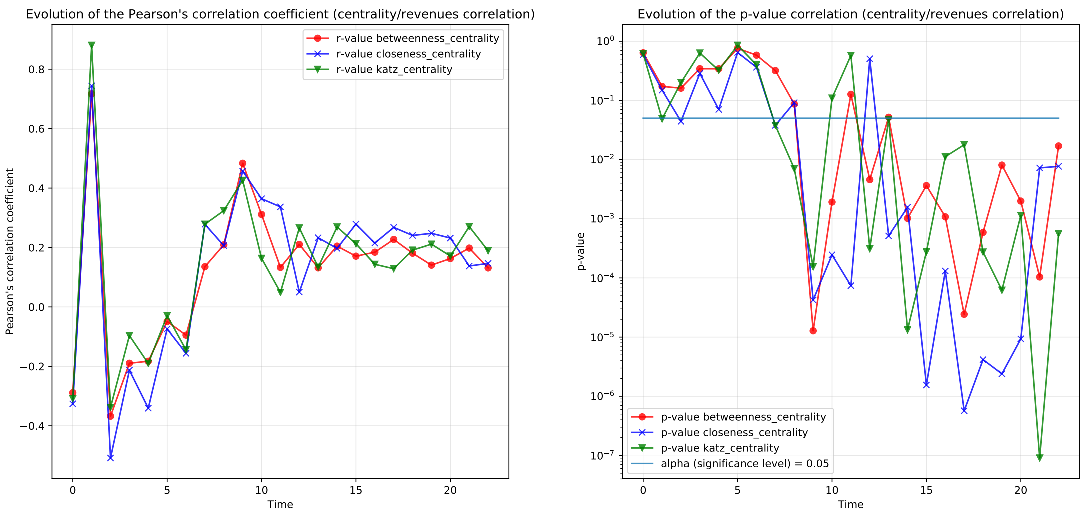
- $Log\ Likelihood\ Ratio\ Test$ for Model 6.1 and 6.5 gives a $p-value$ equal to the value of a Chi-squared distribution of 62, and 3 degrees of freedom.
- $Log\ Likelihood\ Ratio\ Test$ for Model 6.4 and 6.5) gives a $p-value$ equal to the value of a Chi-squared distribution of 19.8 and 1 degrees of freedom.
- $Log\ Likelihood\ Ratio\ Test$ for Model 6.3 and 6.5 gives a smaller $p-value$ equal to the value of a Chi-squared distribution of 13.8 and 1 degrees of freedom.

We could have also used the Akaike Information Criterion (AIC) [23] and the Bayes Information Criterion (BIC) [24], extensively explained in Appendix C 6.11 However, both these two models' selection estimator give results similar to the $Log\ Likelihood\ Ratio\ Test$. Curve in Figure 6.4 shows such results.

Concluding, if we want to go for the best model between all the listed one we should take Model 6.1, for *Adj.R Squared*, $p-values$, $Log-Likelihood\ Ratio\ Test$, AIC and BIC. Otherwise, if we want to reduce the complexity of our model we could just go for Model 6.4, the Katz centrality one. In any case, the notions of centrality seem to correlates with the best business strategy and, thus, legitimize the Client Space approach.

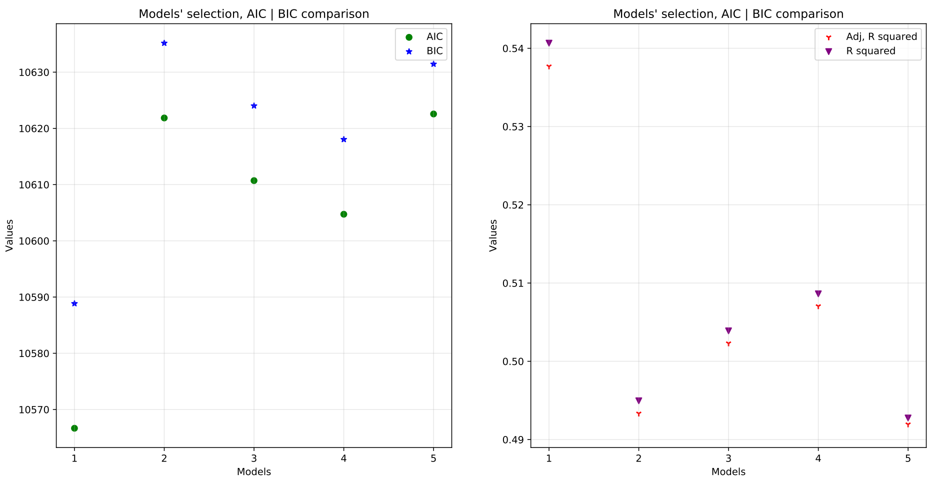
However, the most important results, that need to be underlined, is that we have a higher statistical significance of models that include centrality variables. Hence, proving that positions of customers in the Client Space aren't equally profitable, we have coherently legitimized the introduction of the Client Space as a methodological tool of analysis.

Figure 6.3: Silk Road, Client Space, correlation between centralities and revenues



Evolution of the correlation between vendors' Client Space centralities and vendors' revenues. We correlate revenues with three different metrics of centrality: closeness centrality, betweenness centrality and Katz centrality. The right side of the Figure shows the evolution of the Pearson's correlation coefficient of the tree correlations. Whereas, the left side shows the evolution of the p-values of the three in comparison with the standard significance level, alpha = 0.05. Basically, the left side tells that the three correlations are statistically relevant after the 12th time window, i.e when the p-values start to be consistently smaller than the significance level. Curiously, after the 12th time window, the Client Space starts to reach a stable topology and the Pearson's correlation coefficients start to underline a weak ($r - value \sim 0.2$), but still significant, correlation.

Figure 6.4: Silk Road, model'sselction



Given a unique time window of the second half of the market activity, we are comparing nested models of multiple linear regression between revenues and several parameters, that gradually includes centrality measures and the number of clients variable. The x-axis relates the values to the labels of the models expressed in 6.3. Results on the right side, show how Model 6.1 has the minimum BIC and AIC values, with a significant difference with respect to Model 6.5. Results on the left side, show how Model 6.1 has the minimum Adj. R squared and R squared values, with a significant difference with respect to Model 6.5.

Conclusions

What does determine the success of vendors in Darknet marketplaces?
What are the major factors of competitiveness between vendors?

In this master thesis, we have shown and justified how to cast this question into the scientific grounding of a new economic thinking: the economic complexity approach gave us the opportunity to compare monetary information such as incomes with a measure of the intangibles competitive attribute of each vendor

A rich strategy literature suggests that vendor fitness is driven by organization-specific capabilities to reliably produce specific product sets, which suggests a potential link with the Hildago & Haussman model of capabilities. Yet, the set of products offered by a vendor are typically conceptualized within the tradition of Hidalgo & Haussmann as being reflective of their underlying capabilities, we surprisingly found no evidence for this model in Darknet marketplaces. We suggest that this is one key reason our analysis of the bipartite network failed. The Hildalgo & Haussmann model is sensible at the country-level, as it aggregates the various products made by many independent countries, but is not as sensible at the organization-level, as vendors do not typically benefit from producing a diversity of unrelated products, given poor economies of scale and increased coordination costs to manage multiple unrelated products. Both the the tripartite generalisation of the Fitness-Complexity Algorithm and the parallel analysis of vendors' diffusion over the Client Space and the Product Space help us understanding what determines vendors' competitivenss and what makes it different with respect to country's competitivenss. This new tools brought us to two main findings:

- The first explanation focuses not merely on capabilities to produce specific products, but on dynamic capabilities to update and adjust the range of the product offering set. We found no such evidence that the fittest vendors updated their product sets. Whereas, we found that the fittest and richest vendors all pursued a focus strategy: they focused on a limited subset of products (typically 3-5) while tending their trades to spread over a multitude of clients.
- Furthermore, we saw that these vendors gradually moved from the periphery to the core of the Client Space; namely, the most successful vendors have been able to navigate profitably (probably unconsciously) through the Client Space. Open questions remain: is it something that the vendors pursue actively? Is it a result of the client's talking to one another? We do not know, at least not yet.

At this point, these results raise a spontaneous question:

What is the analogue of the tripartite network of vendors, products and customers in the framework of trades between countries?

We came up with the idea of testing both the metrics of Chapter 4 and Chapter 5 on an Import-Export countries dataset, classifying some countries as "customers" and some others as "vendors". For instance, the 70 % USA' GDP is based on consumption, while, the GDP of China has been mostly fuelled by production and export, idem for Germany and Japan to some degree, Switzerland, Singapore... etc; indeed the world can be considered as divided into two big classes: net exporters and net importers [20]. One could thus improve the Pietronero's approach by subdividing the countries with industry subsections, in which some countries will play the role of the customers of the dark market set-up that has been built along this thesis. We are currently looking for the correct dataset in order to utilize the tripartite metric.

Appendix A

6.4 Betweenness centrality

Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes. In his conception, vertices that have a high probability to occur on a randomly chosen shortest path between two randomly chosen vertices have a high betweenness.

The betweenness of a vertex x in a graph $G := (V, E)$ with V vertices is computed as follows:

For each pair of vertices (s, t), compute the shortest paths between them. For each pair of vertices (s, t), determine the fraction of shortest paths that pass through the vertex in question. Sum this fraction over all pairs of vertices (s,t). More compactly the betweenness can be represented as:[21]

$$C_B(x) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(x)}{\sigma_{st}} \tag{6.6}$$

6.5 Closeness centrality

Closeness centrality of a node is the average length of the shortest path between the node and all other nodes in the graph. Thus the more central a node is, the closer it is to all other nodes.

$$C(x) = \frac{1}{\sum_y d(y, x)} \tag{6.7}$$

where $d(y, x)$ is the distance between vertices x and y. However, when speaking of closeness centrality, people usually refer to its normalized form, generally given by the previous formula multiplied by $N - 1$, where N is the number of nodes in the graph. This adjustment allows comparisons between nodes of graphs of different sizes.

6.6 Katz centrality

Katz centrality is a generalization of degree centrality. Degree centrality measures the number of direct neighbors, and Katz centrality measures the number of all nodes that can be connected through a path, while the contributions of distant nodes are penalized. Mathematically, it is defined as

$$x_i = \sum_{k=1}^{\infty} \sum_{j=1}^{N} \alpha^k (A^k)_{ji} \tag{6.8}$$

where α is an attenuation factor in (0,1).

Appendix B

6.7 Model 6.1

Dep. Variable:	y	**R-squared:**	0.541
Model:	OLS	**Adj. R-squared:**	0.538
Method:	Least Squares	**F-statistic:**	182.1
Date:	Fri, 28 Sep 2018	**Prob (F-statistic):**	4.55e-103
Time:	19:03:19	**Log-Likelihood:**	-5278.3
No. Observations:	624	**AIC:**	1.057e+04
Df Residuals:	619	**BIC:**	1.059e+04
Df Model:	4		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	-6184.1145	1016.526	-6.084	0.000	-8180.373	-4187.856
x1	-7821.8957	2135.904	-3.662	0.000	-1.2e+04	-3627.399
x2	-1.462e+04	3585.872	-4.078	0.000	-2.17e+04	-7581.961
x3	1.791e+05	2.55e+04	7.037	0.000	1.29e+05	2.29e+05
x4	89.5658	4.128	21.699	0.000	81.460	97.672

Omnibus:	589.874	**Durbin-Watson:**	1.708
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	31988.119
Skew:	4.035	**Prob(JB):**	0.00
Kurtosis:	37.135	**Cond. No.**	8.86e+03

6.8 Model 6.2

Dep. Variable:	y	**R-squared:**	0.495
Model:	OLS	**Adj. R-squared:**	0.493
Method:	Least Squares	**F-statistic:**	304.3
Date:	Fri, 28 Sep 2018	**Prob (F-statistic):**	7.68e-93
Time:	19:03:19	**Log-Likelihood:**	-5307.9
No. Observations:	624	**AIC:**	1.062e+04
Df Residuals:	621	**BIC:**	1.064e+04
Df Model:	2		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	-320.2165	62.118	-5.155	0.000	-442.202	-198.230
x1	-2911.8143	1771.767	-1.643	0.101	-6391.195	567.566
x2	91.6863	3.845	23.844	0.000	84.135	99.237

Omnibus:	698.948	**Durbin-Watson:**	1.731
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	67353.668
Skew:	5.141	**Prob(JB):**	0.00
Kurtosis:	52.848	**Cond. No.**	588.

6.9 Model 6.3

Dep. Variable:	y	**R-squared:**	0.504
Model:	OLS	**Adj. R-squared:**	0.502
Method:	Least Squares	**F-statistic:**	315.4
Date:	Fri, 28 Sep 2018	**Prob (F-statistic):**	3.00e-95
Time:	19:03:19	**Log-Likelihood:**	-5302.4
No. Observations:	624	**AIC:**	1.061e+04
Df Residuals:	621	**BIC:**	1.062e+04
Df Model:	2		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	685.8418	285.055	2.406	0.016	126.052	1245.631
x1	-1.266e+04	3390.134	-3.734	0.000	-1.93e+04	-5999.989
x2	97.0017	4.111	23.597	0.000	88.929	105.074

Omnibus:	691.688	**Durbin-Watson:**	1.739
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	64915.927
Skew:	5.056	**Prob(JB):**	0.00
Kurtosis:	51.934	**Cond. No.**	1.14e+03

6.10 Model 6.4

Dep. Variable:	y	**R-squared:**	0.509
Model:	OLS	**Adj. R-squared:**	0.507
Method:	Least Squares	**F-statistic:**	321.4
Date:	Fri, 28 Sep 2018	**Prob (F-statistic):**	1.53e-96
Time:	19:03:19	**Log-Likelihood:**	-5299.4
No. Observations:	624	**AIC:**	1.060e+04
Df Residuals:	621	**BIC:**	1.062e+04
Df Model:	2		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	-4239.7034	869.009	-4.879	0.000	-5946.256	-2533.151
x1	9.756e+04	2.18e+04	4.478	0.000	5.48e+04	1.4e+05
x2	82.0299	3.977	20.627	0.000	74.220	89.839

Omnibus:	649.956	**Durbin-Watson:**	1.711
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	48035.626
Skew:	4.625	**Prob(JB):**	0.00
Kurtosis:	44.976	**Cond. No.**	7.34e+03

6.11 Model 6.5

Dep. Variable:	y	**R-squared:**	0.493
Model:	OLS	**Adj. R-squared:**	0.492
Method:	Least Squares	**F-statistic:**	604.2
Date:	Fri, 28 Sep 2018	**Prob (F-statistic):**	9.57e-94
Time:	19:03:19	**Log-Likelihood:**	-5309.3
No. Observations:	624	**AIC:**	1.062e+04
Df Residuals:	622	**BIC:**	1.063e+04
Df Model:	1		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	-356.5455	58.131	-6.134	0.000	-470.702	-242.389
x1	89.6623	3.648	24.581	0.000	82.499	96.825

Omnibus:	703.006	**Durbin-Watson:**	1.731
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	68911.676
Skew:	5.188	**Prob(JB):**	0.00
Kurtosis:	53.426	**Cond. No.**	19.3

Appendix C

6.12 Likelihood Ratio Test

The Likelihood Ratio Test (LR test) is a statistical test used for comparing the goodness of fit of two statistical models — a null model against an alternative model, [21]. The Likelihood Ratio, 6.9, expresses how many times more likely the data are under one model than the other. Such a value is further used to compute a p-value, or compared to a critical value to decide whether or not to reject the null model. In order to evaluate the p-values, Wilk's theorem plays a major role, [22]. Indeed, if the size of the sample n approaches ∞, the test statistic 6.10 for a nested model will be, asymptotically, a Chi-squared distribution χ^2, with a degree of freedom equal to the difference in dimensionalities of the alternative models and the null model.

$$\Lambda(x) = \frac{L(\theta_0 \mid x)}{L(\theta_1 \mid x)} \tag{6.9}$$

$$-2\log\Lambda(x) \to \chi^2 \qquad if \qquad n \to \infty \tag{6.10}$$

Hence, as in our case, it is possible to compute the likelihood ratio and compare it to the Chi-squared value which corresponds to a specific statistical significance and compared to any standard significance level.

It needs to be noticed that the Likelihood Ratio Test is only valid for nested models, thus applicable in our case.

6.13 Akaike Information Criterion

The Akaike information criterion (AIC), formulated by the statistician Hirotugu Akaike, is an estimator of the relative quality of statistical models for a given set of data, [23]. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection. AIC is founded on information theory. When a statistical model is used to represent the process that generated the data, the model will almost never be exact; so some information will be lost by using the model to represent the process. AIC estimates the relative information lost by a given model in representing the process of data generation: the less information a model loses, the higher the quality of that model. In making an estimate of the information lost, AIC deals with the trade-off between the goodness of fit of the model and the simplicity of the model itself.

$$AIC = 2k - 2ln(\hat{L}) \tag{6.11}$$

With k be the number of estimated parameters in the model and $\hat{L}$ be the maximum value of the likelihood function for the model.

Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Thus, AIC rewards goodness of fit (as assessed by the likelihood function), but it also includes a penalty that is an increasing function of the number of estimated parameters. The penalty discourages overfitting because increasing the number of parameters in the model almost always improves the goodness of the fit. We would then choose the candidate model that minimized the information loss.

6.14 Bayes Information Criterion

The Bayesian information criterion (BIC) is a criterion for model selection among a finite set of models [24]; the model with the lowest BIC is preferred. It is based, in part, on the likelihood function and it is closely related to the Akaike Information Criterion (AIC) 6.13. When fitting models, it is possible to increase the likelihood by adding parameters, but doing so may result in overfitting. Both BIC and AIC attempt to resolve this problem by introducing a penalty term for the number of parameters in the model; the penalty term is larger in BIC ($\log(n)k$) than in AIC ($2k$).

$$BIC = \log(n)k - 2ln(\hat{L}) \tag{6.12}$$

With n be the number of data points, k be the number of estimated parameters and $\hat{L}$ be the maximum value of the likelihood function for the model.
A comparison between BIC and AIC argue that BIC is appropriate for selecting the "true model" from the set of candidate models. To be specific, if the "true model" is in the set of candidates, then BIC will select the "true model" with probability 1, as $n \to \infty$; in contrast, when selection is done via AIC, the probability can be less than 1. Proponents of AIC argue that this issue is negligible because the "true model" is virtually never in the candidate set. Indeed, it is a common aphorism in statistics that "all models are wrong"; hence the "true model" (i.e. reality) cannot be in the candidate set. Hence, they argue that AIC is appropriate enough for finding the best-approximating model, [25]

Bibliography

[1] K. Soska and N. Christin. *"Measuring the Longitudinal Evolution of the Online Anonymous Marketplace Ecosystem."* USENIX Security Symposium. (2015).

[2] N. Christin. *"Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace."* Proceedings of the 22nd international conference on World Wide Web. ACM, (2013).

[3] M. Horton-Eddison and M. Di Cristofaro. *"Hard Interventions and Innovation in Crypto-Drug Markets: The Escrow Example."* GDPO policy brief. Google Scholar. (2017).

[4] N. Lorenzo-Dus and M. Di Cristofaro. *" 'I know this whole market is based on the trust you put in me and I don't take that lightly': Trust, community and discourse in crypto-drug markets."* Discourse and Communication, 1750481318771429. (2018).

[5] L. Armona. *"Measuring the Demand Effects of Formal and Informal Communication: Evidence from Online Markets for Illicit Drugs."* arXiv preprint arXiv:1802.08778, (2018).

[6] M. Horton-Eddison. *"Updating Escrow: Demystifying the CDM multisig process."* GDPO policy brief. Available at:
http://www.swansea.ac.uk/media/HortonEddisonGDPOMultiSigEscrowSA.pdf. (2017).

[7] The Economist. *"Shedding light on the dark web."* Available at:
https://www.economist.com/international/2016/07/16/shedding-light-on-the-dark-web . (2016).

[8] J. Barney. *"Firm resources and sustained competitive advantage."* Journal of management, 17(1), 99-120. (1991).

[9] C. A. Hidalgo et al. *"The product space conditions the development of nations."* Science 317.5837 : 482-487. (2007).

[10] C. A. Hidalgo and R. Hausmann. *"The building blocks of economic complexity."* Proceedings of the national academy of sciences 106.26 : 10570-10575. (2009).

[11] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli and L. Pietronero. *"A new metrics for countries' fitness and products' complexity."* Scientific reports, 2, 723. (2012).

[12] M. Cristelli, A. Gabrielli, A. Tacchella, G. Caldarelli and L. Pietronero. *"Measuring the intangibles: A metrics for the economic complexity of countries and products."* PloS one, 8(8), e70726. (2013).

[13] M. Cristelli, A. Tacchella and L. Pietronero. *"The heterogeneous dynamics of economic complexity."* PloS one, 10(2), e0117174. (2015).

[14] L. Pietronero, M. Cristelli, A. Gabrielli, D. Mazzilli, E. Pugliese, A. Tacchella and A.

Zaccaria. *"Economic Complexity:" Buttarla in caciara" vs a constructive approach."* arXiv preprint arXiv:1709.05272. (2017).

[15] M. Cristelli, A. Tacchella, M. Cader, K. Roster and L. Pietronero. *"On the predictability of growth."* World Bank Document. Policy Research Working Paper 8117. (2017).

[16] E. Pugliese, A. Zaccaria and L. Pietronero. *"On the convergence of the Fitness-Complexity Algorithm."* The European Physical Journal Special Topics, 225(10), 1893-1911. (2016).

[17] M. Cristelli and A. Zaccaria. *"Personal communication with the coauthors of the Fitness-Complexity Algorithm."*

[18] E. Pugliese, L. Napolitano, A. Zaccaria and L. Pietronero. *"Coherent diversification in corporate technological portfolios."* arXiv preprint arXiv:1707.02188. (2017).

[19] V. D. Blondel, J. L. Guillaume, R. Lambiotte, E. and Lefebvre. *"Fast unfolding of communities in large networks."* Journal of statistical mechanics: theory and experiment, 2008(10), P10008. (2008)

[20] Central Intelligence Agency. *"The World Factbook"* Available at: https://www.cia.gov/library/publications/the-world-factbook/rankorder/2087rank.html and https://www.cia.gov/library/publications/the-world-factbook/rankorder/2078rank.html

[21] D. V. Hinkley and D. R. Cox. *"Theoretical statistics."* Chapman and Hall/CRC. (1979).

[22] S. S. Wilks. *"The large-sample distribution of the likelihood ratio for testing composite hypotheses."* The Annals of Mathematical Statistics, 9(1), 60-62. (1938).

[23] H. Akaike. *"Information theory and an extension of the maximum likelihood principle."* In Selected papers of hirotugu akaike (pp. 199-213). Springer, New York, NY. (1998).

[24] E. Wit, E. V. D. Heuvel and J. W. Romeijn. *"All models are wrong...': an introduction to model uncertainty."* Statistica Neerlandica, 66(3), 217-236. (2012).

[25] K. Aho, D. Derryberry and T. Peterson. *"Model selection for ecologists: the worldviews of AIC and BIC."* Ecology, 95(3), 631-636. (2014).