What do interlock *actually* do

Javier Garcia-Bernardo[1, *]

[1] *Department of Political Science, Amsterdam Institute for Social Science Research (AISSR), CORPNET Lab, The University of Amsterdam, 1018 WV Amsterdam.*

(Dated: April 18, 2016)

I. BRIEF DESCRIPTION OF THE PROJECT

The study of corporate networks dates back to the beginning of the 20*th* century [1], when researchers noticed increased relationships between banks and industry. These relationships were created through the appointment of directors with positions in both industries. Nowadays, 25% of the companies worldwide are connected by shared board directors (interlocks). Several explanations (micro-motives) have been pointed out for the creation of such links. The micro-motives include facilitating collusion (cooperate to limit competition), cooptation (absorption of potentially disruptive elements), creating legitimacy by hiring respectable directors, career advancement and social cohesion.

The presence of interlocks have been correlated with macro-outcomes, finding contradictory results with profits, and with the effect in merges and takeover [2]. Only the spread of corporate strategies has been consistently found in the literature [2]. The ambiguous results can be attributed to the biased data studied in the past, namely a few dozen companies in a particular sector. Using data from the Orbis database, comprising 200 million companies and 100 million directors, we will analyze the effects.

Moreover, the fine allows us to study the characteristics of the network without imposing a national character. In this way blablbah (d. In some regions the business communities are organized along national borders, whereas in other areas the locus of organization is at the city level or international level.)

This research proposal is organized as follows. Firstly we explain the characteristics of the data. Secondly we give an overview of complexity methods that can and have been applied to social sciences. Thirdly, we provide an overview of the problems (research questions) within the project that can be tackled using complex systems methods. Finally, we focus on the propositions, concepts and hypotheses of the project.

II. COMPLEXITY THEORY

A. Complex systems

Complex systems are those where many similar parts interact with each other using simple rules to create the whole, which exhibit characteristics different than the parts – “more is different” [3]. Complex systems contrast with complicated systems, where many different parts with defined roles are put together to create the system. Complicated systems, such a watch, can be studied by analyzing their parts. Moreover the failure of a piece produces the failure of the system. Complex systems, such a bird flock, cannot be studying by analyzing only their parts, but the interactions among them are also needed. Although a consensus definition of complex system does not yet exist, a complex system is characterized by the following attributes. 1. Multi-scale: Many individual parts interact to create the whole. 2. Networks: The individuals usually interact with a few other individuals, creating a network of interactions. For social systems, the networks created are “small-world” networks, where the distance between two random people in a network is small. 3. Emergent properties: The whole has properties that none of the individuals have. 4. Spontaneous order. There is not a global organizer of the system. For instance, the standing ovation is an emergent property of the interaction between people. 5. Memory: The individuals remember previous interactions. 6. Feedback loops: The interaction between two individuals affect other individuals in the system. For example your decision about standing up affects the probability that other people start standing up. 7. Stochasticity: The system lives in a noisy environment. 8. Steady-states are far from equilibrium. Complex systems are usually in a steady-state (except during transition times). However, since they depends on active interactions between people, they stay far from the equilibrium. If no energy is added to the system, the system disappears. 9. Nonlinearity, cascading and hysteresis. The interactions are non additive. For the standing ovation, the addition of a new person standing can produce a cascade of events. 10. Robustness to random failures. The system is highly resistant to the failure of one of the individuals. 11. Sensitivity to targeted failures if individuals are organized in network: The system is sensitive to the failure/removal of a few specific individuals.

* garcia@uva.nl

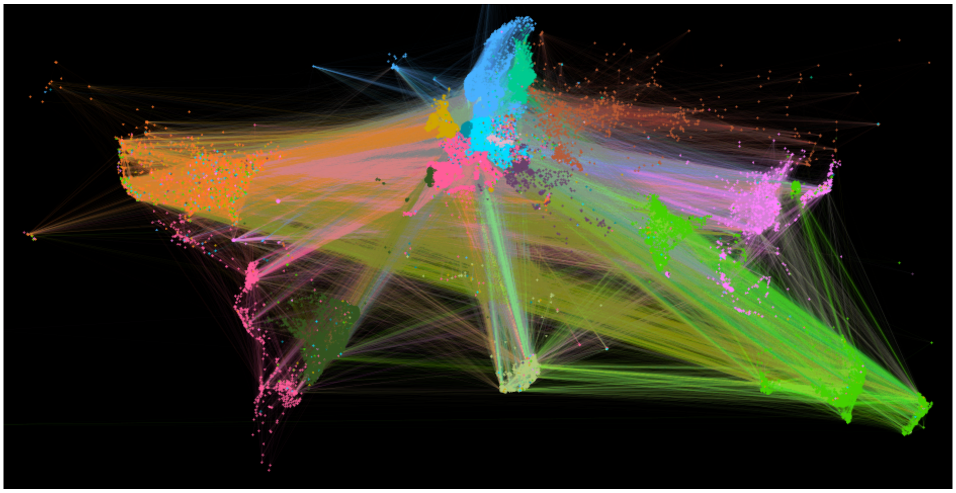

FIG. 1. Global network of interlocking directorates. Color indicates communities – i.e. cities that do business together within each other more often than with others.

B. The micro-macro problem

Some social systems exhibit different properties than the parts composing them. In those systems, the action of an agent cannot be solely predicted from its characteristics, but the interactions between agents need to be taken into account. For instance, it is more likely that you start using purple hats if your partner thinks that are trendy than if your colleague does, which in turn affect the probability of your friends and colleagues also wearing purple hats. The probability that every person in society start using purple hats cannot be predicted by their perceived trendiness, but networks must be taken into account. Similarly, the performance of a firm depends not only on their product, but also on the interaction between firms, governments and other groups, which in turn are composed and influenced by individuals. Moreover, the success of their product – which is partly based on the interaction with other firms and groups – affects the actions of your suppliers and competitors. A classic example are format wars, where inferior products can (and often do) succeed. The problem where the whole depends on both the parts and the interaction between pars is often called 'the micro-macro problem'.

The micro-macro problem is found across disciplines. In physics, snowflakes form by the interaction between low-energy H_20 molecules, where the specific shape of the flake depends on the interactions between the individual molecules. In biology, organs are composed of individual cells that do not have the properties of the organ. A cardiac cell alone lack the capability to beat, however a few hundreds of cells together spontaneously start to beat. In ecology, ant colonies are efficiently organized to collect food, clean and defend the colony, and reproduce. However an individual ant cannot perform all those tasks, but requires chemical stimuli from other ants to coordinate. In social systems, complex behaviours are the results of the interactions between the agents. A mediocre play can receive a standing ovation if a group stands up immediately after the play ends, creating a cascade of people standing up [4].

Importantly, although predicting the whole in individual situations is difficult, this does not imply that we cannot observe correlations at the macro scale. In our physics example, humidity and temperature increase the probability of a specific shape of the snowflake. Cardiac cells will be more likely to beat if specific chemicals are added. A colony of ants will be more likely to leave the colony to scavenge for food if the temperature is low. A play will be more likely to receive a standing ovation example if it is good. However, it is important to observe that the correlations are indirect. The humidity does not affect the shape directly, but affect the probability of two molecules of water binding in an specific way. The quality of the play will increase the probability that some people stands up, creating a cascade of people standing. Because macro outcomes are non-linear aggregations of the micro agents, an appropriate model for the micro-macro problem is required.

C. The micro-macro problem in social sciences. A complex systems perspective

In social sicences, Collemans scheme [6] (Fig. 2A) is the standard framework to represent the micro-macro problem. Nodes A and D in Figure. 2 are the macro-conditions (composed of the environment where the system is situated) and macro-outcomes of the social system. Node B corresponds to the micro-conditions (composed of the perceived environment, as well as genetic and other individual factors). Micro-conditions are affected by

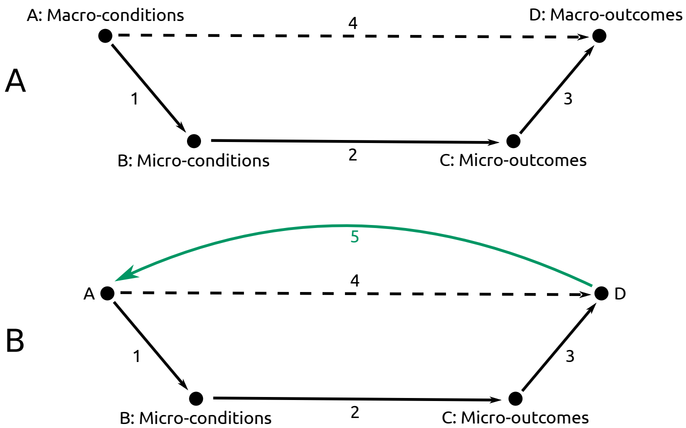

FIG. 2. Global network of interlocking directorates. Color indicates communities – i.e. cities that do business together within each other more often than with others. Adapted from [5]

the macro-conditions. This arrow is usually labeled as bridge assumptions. Micro-outcomes are the decisions of the individual among the possible options. In the standing ovation example, this corresponds to each individual decision to stand up or not. This arrow is non-trivial, since the decision does not depend only on each agent's conditions, but also on interdependent relationships with other individuals. In Granovetter's model [7], the individuals are characterized by a threshold ϕ that summarizes their micro-conditions. Each person will then riot if there are at least ϕ other people rioting already. Similarly to the standing ovation, the population will end up in a generalized revolution, or it will die off depending on the distribution of thresholds in the population. For instance, consider population A, where 10% of a population starts rioting, but the other 90% will not riot unless 20% of the population is rioting. In this population, the 'average' person will riot if 18% of the population are rioting, but no revolution will occur. Imagine now population B, where 10% of the population starts rioting, another 20% will riot if 10% is already rioting and the final 70% will only riot if 30% of the population is rioting the revolution will spread. In population B, the 'average' person will riot if 23% are already rioting, however due to the non-linearity in the aggregation a revolution will occur. The aggregation of the micro-outcomes (the individual decisions to riot or not riot) corresponds to label 3. This arrow is usually labeled as transformation rules.

Colemans original scheme [6] uses Webbers origin of capitalism [8] as an example, linking it to the micro-macro problem. He explains the rise of capitalism (D) from protestant religious doctrine (A). A protestant religious doctrine creates specific values in the individuals (B) that produce certain economic behaviours (C). The aggregation of these economic behaviours gives rise to the capitalism (D). Capitalism is the result of the economic behaviour of people, which in turn is caused by the individual interpretation of the values of protestantism. A revolution is the result of the decision of people to riot given their angriness level caused by the macro-conditions. While apparently coherent, this reasoning has two main flaws, Firstly, it obviates the link from (D) to (A). The origin of capitalism was a process that lasted decades, where the economic behaviours produced some intermediate macro-outcomes that affected the macro-conditions. This in turn affects the micro-conditions (values), which affect the micro-outcomes (economic behaviour) from the previous time point. Capitalism is one of the steady states of the cycle. Our view of the process, using Granovetter model as an example, is summarized in Figure 2B. In Granovetter example there are some specific macro-conditions at time zero, such as a level of hunger, a level of police reprisal, or a sense of collective. Moreover, no people are rioting ($\Phi_0 = 0$, where Φ_0 is the percentage of the population rioting at time 0). Given the micro-conditions, a few people (x) with threshold zero ($\phi_i == 0 \geq Phi_0$) – very prone to riot – make the individual decision to riot (micro-outcome). The macro-outcome is that a group of people are rioting. This affects the macro-conditions ($\Phi_1 = x$). The micro-conditions in Granovetter model are fixed (constant thresholds, although this is a simplification). Now, every individual compares their own threshold with Φ_1, and riot if $\phi_i \geq Phi_1$. The cycle continues until we reach a stable state (for example revolution or just a few people rioting). This class of systems where the system is continuously evolving correspond to Complex Adaptive Systems.

The second flaw is that it creates an illusion of determinism. In any complex system, random events can create a cascade of events that cannot be explained solely by the macro-conditions. Furthermore, we can always find a logical explanation of the end result. For instance, we can deduct that that few people started rioting because police reprisal, and that produced the cascade. However these kind of 'ad hoc' explanations makes police reprisal a necessary condition, while the same could have happened with no police reprisal, or no rioting could have happened with the same reprisal. While we could probably say that some degree of police reprisal increased the propensity of rioting, we cannot conclude that reprisal is a necessary cause of rioting, or that police reprisal will cause another revolution in similar conditions. When the end result is the non-linear aggregation of many interconnected actors, the results can only be generalized when we can compare many independent, similar cases [9].

In the past decades we have seen the recollection of large datasets for many complex systems. These large dataset include vast information on biology (e.g. interactions between thousands of proteins in the cell), social interactions (e.g. social networks, movement trackers, etc), or the economy (e.g. Orbis, Reuters or LexisNexis provide information on firm indicators and directors for millions of companies worldwide). As a consequence, new methods and models have been developed for the study of these rich datasets. These tools can be grouped in three categories. 1. Descriptive tools: To characterize the macro-outcomes and find patterns in the data. 2. Generative modelling: To explain emergence of macro-outcomes from the micro scale, how the macro-conditions affect the micro scale, or both. 3. Predictive tools: A model is useful if it provide insights in the causal mechanisms and can predict future events. This include predicting what has already happened using only a portion of the data. An example that combines the three categories is the work that resulted in the Atlas of Economic Complexity and that will be the basis of next section [10–13]. Firstly, they described every country (macro scale) with respect to the type and amount of exports and imports (micro scale). Secondly, they assumed a series of capabilities required to produce a product (for example institutions, materials, human capital). If countries produce the products when they acquire the capabilities required to produce it, then products that are often exported together will require similar capabilities. Since it is not possible to quantify those capabilities, they created the one-mode projection of the model, namely network of products (named the ‘product space’), where two products were closer in the network if they were usually exported together. Thirdly, they modeled the development of countries, showing that countries develop by acquiring new capabilities and producing products that are close in the product space. For instance, a country may develop an electronic industry if they chemicals is an important export, but not if they rely on exporting cereals [11]. Finally, they used the current products produced in the country to create a better indicator for the economic growth of the country [13]. Linking back to Colemans modified scheme (Fig. 2B) we see how the products that a country currently export and import (macro-conditions at time zero) affect the products that companies can produce (micro-conditions at time zero). This in turn affect the products that they actually produce (micro-outcomes at time zero), which results in the total exports and imports (macro-outcomes at time zero). In this model, the macro-outcomes at time zero correspond to the macro-conditions at time one. Understanding the cycle has utility to focus investment on certain sectors.

D. Complex system toolbox for network analysis

We next (very) briefly summarize general methods to study data in complex systems. Because most social and economical systems are embedded in networks, we will focus on complex networks. The standard notation treats networks (graphs) as a series of nodes and edges, where nodes are the agents and edges the interactions between agents.

1. Descriptive tools

Descriptive tools are used to find patterns in the data. Summary statistics, correlations and visualizations allow for a rapid exploration of the data, and an assessment of data quality. Two of the main tools are community detection and centrality analysis. Community detection finds nodes that interact with each other more than with the rest of the network. For example in a recent paper we studied a network of firm interlocks (ref), where the nodes correspond to firms and the edges to shared directors. We showed that in some regions the business communities are organized along national borders, whereas in other areas the locus of organization is at the city level or international level. Centrality analysis grades the importance of a node in the network. Different centrality measures have been developed. For instance in betweenness centrality a node is important if it connects far regions in the network. The specific measure will depend on the problem. If we are measuring the spread of information, as Granovetter did in the strength of weak ties [14], betweenness centrality would be indicated.

Communities and centralities have been used abundantly to describe data. In Clauset et. al. [15] description of inequality in academia, they mapped the academic trajectory of 19,000 faculty in three disciplines. From the mapping, they rank each institution by its relative “prestige centrality”. Institution A is more prestigious than B if people can do their PhD in A and move to B, but PhD graduates from B cannot find a job in A. They showed that the top 25% of the institutions positioned 71-86% of all tenure-track faculty, revealing steep inequality and clear hierarchical networks.

2. Modeling

Modeling allows us to understand the world and predict future events better than people [16] . For example, modeling the network of bank co-risk (Fig. 3) allowed authorities to assess the risk of not bailing out banks. The end result was the bankrupcy of Lehman Brothers and the bailing out of AIG. The goals of modeling are to make causal inferences about a phenomenon and to predict future events. Three main models are used: Firstly, traditional statistical models explains an independent variable in terms of dependent variables. They are characterized by a formula whose parameters are estimated. The parameters usually have a direct real-life interpretation. They emphasize inference and work well when the number of dependent variables is small. The most common example of the group is regression. Secondly, machine learning are also used to explain the results of an output (our dependent variable) in terms of an input (independent variables). Opposing statistical models, they are usually black-boxes, meaning that real-life interpretation of the weights do not exist. Consequently, they emphasize prediction and are used when the number of dependent variables is large (big data). Finally, mathematical and computational modeling allows to recreate the physical system and understand the causes that drive the output of the system.

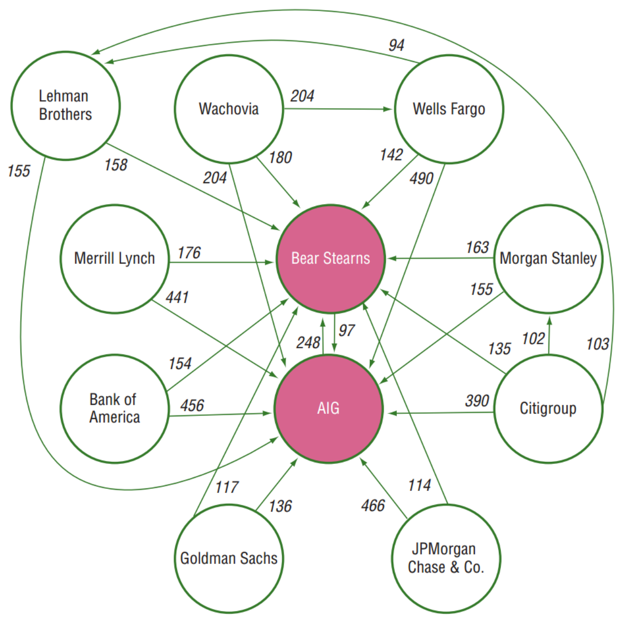

FIG. 3. Sources: Bloomberg, L.P.; Primark Datastream; and IMF staff estimates. Note: This figure presents the conditional co-risk estimates between pairs of selected financial institutions. Only co-risk estimates above or equal to 90 percent are depicted. See Table 2.6 for further information. [17]

Mathematical and computational modeling

As previously discussed, it is not possible to explain emergent properties just by studying the agents. Mathematical and computational models allow to close the gap between the micro and macro scales. In Granovetter rioting example, a mathematical model allows us to understand why a population will end up rioting, while other more prone to riot on average will not. In the segregation example of Schelling [18], a neighborhood of houses is simulated in a square lattice (chess board), where most houses are occupied but some are empty, allowing the inhabitants to move between houses. Two different types of person live in the houses (lets call them Belgians and Dutchies) in equal proportion. None of them are racists, but they would like to have at least two neighboors that are like them. If they do not have at least 25% of the neighbors that are like them they move at random to one of the empty houses. While the at random will not hold in reality, it is a conservative scenario. If segregation is found with this simple model, selectively moving to neighbourhoods with a large population of your kind will only produce higher segregation rates. The result is the clear segregation showed in Figure 4A , that resembled the segregation in life (Fig. 4).

Mathematical and computational models can help us discover how simple rules can produce complex macro results. Models can be classified in several categories. According to their assortativity, they are classified into perfect mixing models, where all the agents can interact with all other agents, and network models, where the agents can only interact with a subset of other agents. According to the presence of noise they can be classified into deterministic and stochastic, where some amount of noise is included.

According to the method used to solve the model, they can be classified into equation-based models (analytical and numeric), where the population is represented with an equation, and agent-based models, where each individual is modeled individually (see Epstein [19] for an excellent review and implications of agent-based models in social sciences). The type of model will depend on the application. For example if we want to measure the effect of time and distance on terrorist attacks, using an analytical model (Hawkes process), where an attack produces a cascade of events, will be useful [20]. If you are interested on modeling smoking behavior in high schools, an agent-based model that includes selection and influence terms may be more appropriate [21].

In political sciences, such models have been successfully applied to understand the spreading of ideas (contagion such as granovveter, Generalized contagion models (Dodds and Wattas 2004,2005): memory of exposure to a contagious entity (e.g. a rumor or disease), variable magnitudes of exposure (dose sizes), and heterogeneity in the susceptibility of individuals), to understand the factors affecting polarization vs homogenization. to find the micro-motives in the origin of inequality, show why coordination and collaboration emerge, among others (refs for all). In general, modeling allows us to find and quantify which micro and macro-conditions are associated to the macro-outcomes observed.

3. *Prediction*

Modeling alone .

For this reason, prediction capabilities are often required.

Another way to

null models

Epstein reference

E. Summing up

Social science systems are often embedded in complex networks of interconnected agents, where the action of an agent affect the actions of the others in a non-linear fashion. With the advances in technology, we are for the first time able to gather large datasets of social systems. Data availability has come together intensive research in the area of complex systems, and we have now models to describe, model and predict these challenging structures.

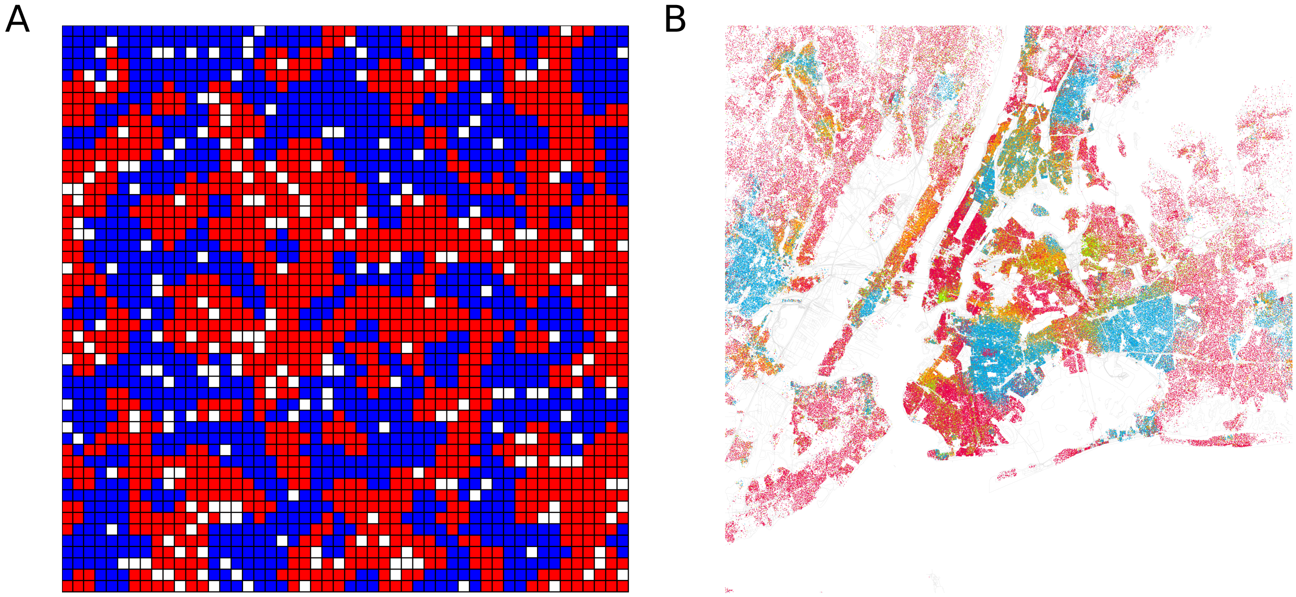

FIG. 4. Gwhite is pink (heavier concentrations look reddish), black is blue, Hispanic is orange, and Asian is green..

III. RESEARCH QUESTIONS, CONCEPTS AND PROPOSITIONS

A. Interlocking directorates

An important concept in this paper is 'interlocking directorates'. An interlock corresponds to the link that is formed between two companies when a director affiliated with one organization sits on the board of directors of another organization [2], or when two director from two firms sit together on the board of directors of a third firm. Importantly, while the interlock is a relationship between companies, it is created by individuals sitting in boards. This allows us to model it as a bipartite (two-mode) network, and to analyze at the same time the importance of both companies and directors.

The concept of interlocking directorates is related to the corporate elite, part of the power elite. [22] defined the power elite as *"those political, economic, and military circles, which as an intricate set of overlapping small but dominant groups share decisions having at least national consequences. Insofar as national events are decided, the power elite are those who decide them"* [22]. Moreover, Mills determined that there is an 'inner core' of the power elite involving individuals that are able to move from one seat of institutional power to another. In the context of corporations, this inner core corresponds to the corporate elite. A strength of using the concept of interlocking directorates as opposed to corporate elites is that we do not assume a priori characteristics of the elite. Corporate elites on the interlock network are the actors with high centralities values, as measured by network algorithms.

The concept of interlock comes not without problems. We are measuring relationships between companies. However it is not clear that formal interlocks – those occurring by people sitting together on boards – are more important than interlocks created by directors being part of the same social clubs or the same families. By restricting ourselves to formal interlocks we may be missing an important part of the corporate network. However, we expect to capture a significant part of the relations between companies.

The effects of interlocks (or by extension the corporate elites) are still object to controversy. In the past decades, tens of papers have investigated the effect of interlocks in firm performance, innovation, acquisitions and mergers, capital growth, firm reputation, and adoption of structures and strategies (see review marzi, x). However, only the spread of structures and strategies has been consistently associated to interlocks (refs). We attribute this to bias to data collection. Previous papers have focuses on a small number of top companies (10–1000), many times focusing only on one sector or country. The result of this may be a misrepresentation of the patterns. For instance, [2] points out that if directors are appointed when the productivity of the firm is low (as a form of monitoring), and also decide to join boards that are performing well (as a form of career development), we could explain the inverted U shaped relationship between number of interlocks and firm performance.

We will study a large dataset comprising 200 million companies and 100 million directors (see IV) to bring definite answers to the field. Our main research question is "How is the interlocking directorates network structured in time and space, and what is their effect on investment and innovation". For the sake of brevity, we will focus here on the second part of the question "what is their effect on investment and innovation". First, we will explore the concept of the product-service space (PSS), that determines how closely two sectors are depending on how many companies in our unit of analysis. Secondly, we will determine how interlocks affect the PSS. Thirdly, we will show what factors influence the presence of interlocks. Our unit of analysis can be the company itself, the city, the country or the region. For clarity we will focus on the city level for the rest of the manuscript, although we may do the analysis at several scales.

B. Product-service space

[Research subquestion herere] Cities and countries develop economically by moving from producing simple products and services to specializing in more expensive ones – a process referred to as 'structural transformation' [11, 23–25]. This transformation can be explained using the relative proportion between productive factors and technological differences (see [13] for a review). Current models usually abstract from the product themselves and look at macro-economic indicators of productive factors and technological differences. However, growth occur when individual new products are created or existing ones improved, and it is not clear if . Moreover, the products that are developed depend on the current products being produced – there is a relatedness between products. Many explanation have been proposed, such as similar institutions, infrastructure, physical factors, technology, or some combination of those factors (see [11]). For instance, [11] points out that making cotton shirts does not require more or less skills than making chocolate, but different skills. Knowing what sectors a city or a country has a competitive advantage can allow us to predict next.

[10–13] developed the 'product space' as a map of the relatedness between products, where products are related if they are produced in the same country. They showed that structural transformation at the country level occurs by moving from existing products to related products, where two products are related if they are closed in the 'product space'. The product space capture information about the complexity of the set of capabilities available in a country, is strongly correlated with income per capita and predictive of future growth. The product space is defined by using an adapted Revealed Comparative Advantage. See [11] for a description on how to build the product space. We will define the network as $PSS = (V, E))$, where V are the sectors according to NACE rev. 2 (e.g. C11 = Manufacture of beverages), and E is the matrix of weights between sectors.

We adapt their concept by focusing at relationships between company sector at the city level, instead of product category at the country level. Using cities allows us to explore not only products, but also services, thus we define the 'product-service space' (PSS). In a world full of multi-nationals, the innovation is happening at the city level [26]. Not, and a better understanding of the situation.

Moreover, focusing on city allows to not care about changing size.

Development is understood as adquiring the underlying capabilities that allow for the diffusion process.

By focusing on the diffusion process instead of increase of gdp we (via the obvious possible reverse causality of how economically thriving cities ought to foster more interlocks than less thriving cities)

C. Interlocks and the PSS

[Research subquestion herere] We will analyze if interlocks are a predictor of diffusion between sectors in cities. There is a relationship between development and the presence of interlocks. Companies situated close geographically, or in places with similar language or colonial ties have greater chances to interlock (refs). Since the establishment of companies in the city allow for greater possibilities of interlocks, we expect a relationship between the PSS and the number of interlocks between sectors. However it is not clear if interlocks are only a cause of development but also an effect. Interlocks provide a communication channel between companies, and serve as a link for the spread of strategies and structures (refs). For instance, [examples of previous research]. We hypothesize that interlocks serve as a communication channel for opportunities, thus increasing investment and R&D to sectors close in the PSS. The increased investment and innovation has been linked to development [11, 24, 25]. In a first step, we will test if interlocks affect the diffusion process in the PSS. In a second step, we will analyze collaboration between companies using patent data to show if this diffusion process is mediated (at least partially) by innovation.

1. *Interlocks affect the PSS*

[Research subquestion herere] Figure 5 shows our approach. Given the number of companies in city A at time t (Fig. 5A), we want to explain the evolution to time t+1 (Fig. 5B). [12] showed that this evolution is related to the 'product space' (Fig. 5C). We can create a network of interlocks (Fig. 5C), where two sectors are connected if a director sits in companies from both sectors. Importantly, the network of interlocks have relationships between sectors that are not present in Figure 5A. This because interlocks are not restricted to the city itself. A director can sit in the banking sector in city A and in the IT sector in city B, even if there are no IT companies in city A.

One method to study if interlocks are predictors of the diffusion process is conditional entropy $H(X|Y)$. The conditional entropy $H(Companies_{t_1}|Companies_{t_0})$ quantifies how much information we need to define the structure of the network at time 1 knowing

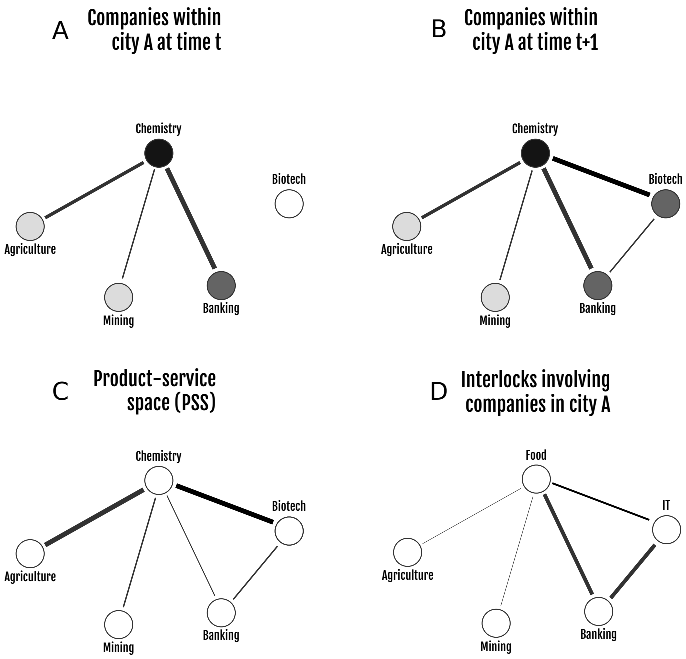

FIG. 5. .

the structure at time 0. If the network of interlocks between sectors affect the network we will find that $H(Companies_{t_1}|Companies_{t_0}, Interlocks_{t_0}) < H(Companies_{t_1}|Companies_{t_0})$. We will compare with random models where the profit and others is included.

There are several problems with this approach. Firstly, we have explained that development creates interlocks (endogeneity problem). The use of longitudinal data allows us to quantify the effect of interlocks on development, independently of the effect of development on interlocks. Secondly, there is a chance of self-selection bias. Companies that want to develop to a new sector may create interlocks with that sector beforehand. However, if this is true, interlocks would also facilitate diffusion. Finally, a most important bias is omitted variable bias. If there is an underlying mechanism that produces the interlock at $time_1$ and the diffusion at $time_2$, we would find a false effect of interlocks on the diffusion process. In order to investigate this possibility, we need to control for city economic indicators, such as infrastructure, resources, education, population density. Other variables to control are sector size, number of companies, and country indicators. Comparing the diffusion process in cities where interlocks were involved and where they were not involved can also help bring insight to this issue. It is important to realize the presence of the , and make sure to detail them in the case the method works.

Rewiring.

2. *Interlocks affect the company space (at least) through innovation*

The subquestion corresponding to this subsection is "Do interlocks foster collaboration between companies and innovation?". In the case of ,

D. Which factors affect interlocks

The literature for the factors influencing interlock creation is more consistent. Geographical distance, colony history, language, education and social networks have been pointed as factors influencing the presence of interlocks (ref). Time allowing, we will explore this idea at the micro-scale using generative models such as ERGMs, which allow to control for network effects such as transitivity – i.e. the probability that there is an interlock A–C if there are interlocks A–B and B–C is increased, and should be controlled.

E. Other projects

The data allow for the exploration of other projects, such as: (i) Description of inequality. (ii) Homogenization of coordinated and liberal market economies. (iii) Quantify the independence of a given sector (e.g. food). (iv) Measure the transference of power from domestic corporation to transnational corporations. (v) Network motifs, which combination of interlocks between sectors are more likely than random.

IV. DATA

"The most satisfactory sampling design for structural analysis is a saturation sample of the entire universe or population; however, this alternative is clearly not feasible for large social structures"

— *M.P. Allen, 1974* [27]

A. Data description

The data from this project was extracted from Orbis. Orbis contains standardized information about 200 million firms and 100 million people. The firm data includes economic indicators (such as turnover, employee number, profit ratios or sector), as well as 90 million ownership ties. The directors data includes biographic information (such as name, education, nationality or gender), as well as 151 million position information, indirectly creating 1,000 million interlocks.

The first step of the project consists on downloading, structuring and storing the data (months 1-4), quantifying the quality of the data (months 4-7) and assessing the effect on the results (months 8-9). For brevity we will skip most the details of the quality assessment. Figure 6A compares the number of companies in Orbis with the OECD. We can see that the quality is extremely good for large companies, but relatively bad for small companies. Importantly, the companies in Figure 6A are those with revenue and employment information (60 million instead of 200). This is, most companies are in the database, but without financial information.

We developed a two-step approach to assess the quality of the data. In the first step, we developed interactive visualizations to rapidly explore the data [28]. The results showed that richer countries (measured by GDP per capita) have larger companies and better coverage. The visualizations also showed that the observed average revenue for the companies in a country depends on the coverage. Those with higher coverage include also small companies, decreasing the 'average' company. In the second step, we characterized the data and extrapolate the quality to other countries. The distributions of revenue for a country follow a lognormal distribution, thus can be defined using two parameters (loc and scale). Moreover, the scale is fixed for all countries and the loc can be estimated using macro-economic indicators. This allows to quantify the type of missing companies (Fig. 6B). In the next month I will assess the effect of this in network measures.

B. Data bias

In terms of internal validity, the concept of interlocks is captured in the data when one director holds positions in two companies. There are some problems with internal validity. Firstly, its not clear that we can separate administrative ties (mainly within corporate groups) from directorate ties (across corporate groups). This is so because the data providers from some countries (including the US) do not always state the type of position that a person holds.

External validity and reliability are intimately related in my case. Since Im studying global data, I can generalize only if my data has high quality. If my data is good everybody can use the same methods and obtain the same results.[they can do that also if the data is bad, but of course you are right to want good data; you should look into the concepts of (internal) validity, reliability, and generalizability again you have not fully understood them]

Instead of talking about validity and reliability I will talk about completeness (how much of the data we have), bias (is our sample a random sample) and accuracy (is the data we have true). They are related to the concepts of validity and reliability and I believe they are more helpful in my discussion. I have extracted data about companies, with economic and geographic information about the company; and directors data, with information about people and the positions they hold.

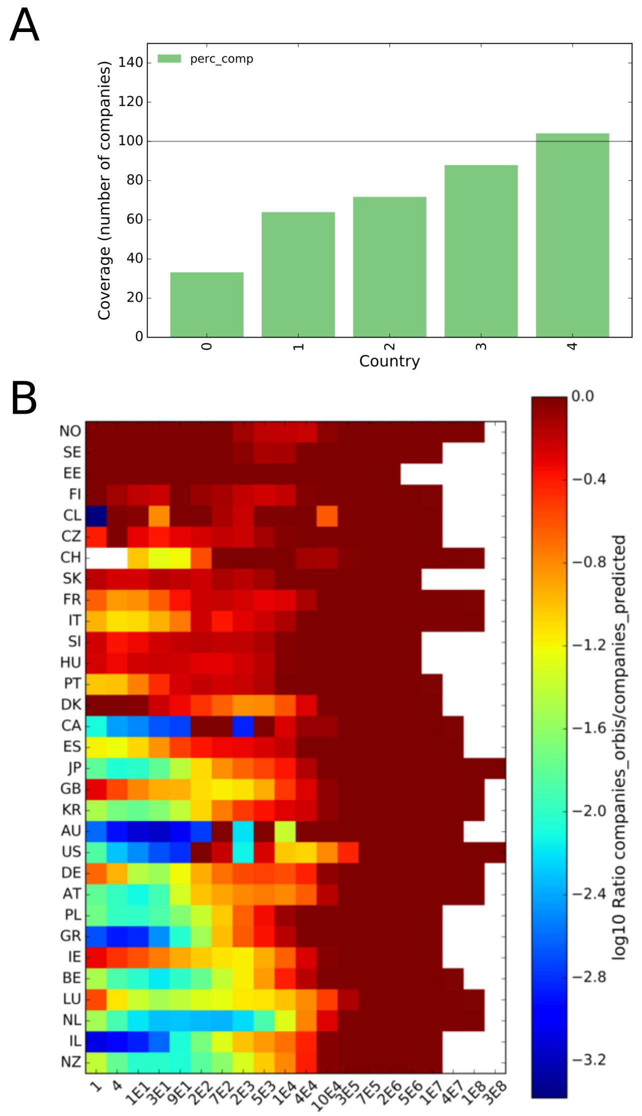

FIG. 6. Global network of interlocking directorates. Color indicates communities – i.e. cities that do business together within each other more often than with others.

Company data 1. Bias and completeness: We have assessed the bias and completeness of the data. We compared Orbis data with OECD data and found that small companies are underrepresented but large companies are always present. Moreover, we developed a method to use the economic data associated to each company to extrapolate the completeness to other countries. In general, Scandinavian countries have the highest quality (very complete and fairly unbiased), while poor countries have very bad data, although their biggest companies are present. We can now use our method to reconstruct missing data and assess the effect of including it. We expect the effect to be small since small companies are not usually connected in the network – for example the owner of a small shop is not likely to sit in the board of directors of any company. [this means that there is bias, but you can assess the direction of the bias: it will lead you to find more interlockedness than there really is; this is important information; it is a problem if you want to make the case that there is a lot of interlockedness, because then you are facing a situation in which your data is biased in a way that helps you prove yourself right; in that case it is better to propose that interlockedness is low, and then be able to show that it is still high, even if small (non-interlocked) companies are underrepresented in the data]

(this is actually the first paper of my PhD. I chose to come up with a new idea for the tasks since I believe I can learn more from the discussion in a more social science paper).

2. Accuracy: We have checked the accuracy of a random sample and the information in the database is almost always correct. However, there are some problems in this area. Some information providers send their data faster than others, which makes the information of some countries be outdated (usually up to one year). Moreover, large company generally are required to fill consolidated accounts, including in their reports the profits, employees and other economics of their subsidiaries. This doubles the information for big companies. [this is not clearly written]

We have found a way to use the ownership relationship between companies to unconsolidate accounts. [good] For the companies with over 250 employees, we now have around 100% of the expected numbers according to OECD data, while before we were almost doubling the numbers.[this is not clearly written]

Directors data 1. Bias and completeness: Based on manual inspection of a small sample of the data we found that some directors from small companies are missing. However, we will try to impute this data based on other economic factors (work in progress).[how is that possible?]

2. Accuracy: The accuracy of the data is not great either. Large companies have extra directors that have already left the company. The position of directors in some countries is unspecified, meaning that we dont know if they are administrative ties or not. We can remove administrative ties by using the ownership database. If a director has positions in a shareholder and a wholly owned subsidiary then we can filter that tie. The problem with this approach is that two subsidiaries that are wholly owned by company A may not have any ownership relationship between them, but they are still the same corporation. We need to think more about this problem and try different solutions.[related to my comment above, this means you are adapting your data pool to be able to stick to your original definition of interlocks]

In general, the way to assess validity and reliability in my paper is to assure good data quality. [data quality is only one source of validity; there are many more issues to consider to maximize the validity of your conclusions] We will do this by imputing the missing data and using this to produce confidence intervals of our metrics. Assessing data quality correctly, and analyzing the entire global network (instead of the top X firms) may allow us to give definite answers to many questions. [your approach to your research is very much data-driven; while you are right to be excited about your data, as it is a rich source of information and can help you make many interesting claims, it is important to get your head out of the data, so to speak, and think strategically about the questions you want answered and the way to answer them]

In terms of operationalization, it is not clear that we can distinguish interlocking directorates from administrative ties. While the first are created by people with ability to decide or control decisions, in the second case the two ‘companies are not independent, but are part of the same structure. Our data does not specify for some countries if the person creating the relationship is a director – and therefore the relationship is a interlock – or if the person is an administrative. However we expect to be able to correct this by using ownership information about companies. Relationships between companies that are accompanied by an ownership relationship can be discarded.

V. SCHEDULE

1. Year 1:
   - Months 1-4: Get data
   - Months 4-7: Analyze quality and bias
   - Months 8-12:
     - Effect of quality
     - Inequality in last names
2. Year 2:
   - Create PSS
   - Analyze effect PSS and city development
   - Descriptive interlock
3. Year 3:
   - Interlocks and PSS
   - Interlocks and innovation
4. Year 4:
   - Determinants of interlocks
   - Write thesis

[1] V. I. Lenin. *Imperialism: the highest stage of capitalism*, volume 1. 1999.

[2] M. S. Mizruchi. What Do Interlocks Do? An Analysis, Critique, and Assessment of Research on Interlocking Directorates. *Annual Review of Sociology*, 22(1):271–298, 1996.

[3] P. W. Anderson et al. More is different. *Science*, 177(4047):393–396, 1972.

[4] J. H. Miller and S. E. Page. The standing ovation problem. *Complexity*, 9(5):8–16, 2004.

[5] W. Raub, V. Buskens, and M. A. Van Assen. Micro-macro links and microfoundations in sociology. *The Journal of Mathematical Sociology*, 35(1-3):1–25, 2011.

[6] J. S. Coleman and J. S. Coleman. *Foundations of social theory*. Harvard university press, 1990.

[7] M. Granovetter. Threshold models of collective behavior. *American journal of sociology*, pages 1420–1443, 1978.

[8] M. Weber. Die protestantische ethik und der geist des kapitalismus. 1904.

[9] D. J. Watts. *The collective dynamics of belief*. Stanford University Press, 2007.

[10] R. Hausmann and B. Klinger. Structural Transformation and Patterns of Comparative Advantage in the Product Space Ricardo. *Faculty Research Working Papers Series Structural*, 2006.

[11] C. A. Hidalgo, B. Klinger, A.-L. Barabasi, and R. Hausmann. The Product Space Conditions the Development of Nations. *Science*, 317(5837):482–487, 2007.

[12] C. A. Hidalgo and R. Hausmann. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26):10570–10575, jun 2009.

[13] R. Hausmann and C. A. Hidalgo. The network structure of economic output. *Journal of Economic Growth*, 16(4):309–342, 2011.

[14] M. Granovetter. The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[15] A. Clauset, S. Arbesman, and D. B. Larremore. Systematic inequality and hierarchy in faculty hiring networks. *Science Advances*, 1(1):e1400005–e1400005, 2015.

[16] P. Tetlock. *Expert political judgment: How good is it? How can we know?* Princeton University Press, 2005.

[17] I. Währungsfonds. Global financial stability report–responding to the financial crisis and measuring systemic risk. *April, Washington, DC*, 2009.

[18] T. C. Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 2006.

[19] J. M. Epstein. Agent-Based Computational Models and Generative Social Science. *Generative Social Science*, 4(5):4–27, 2006.

[20] J. Garcia-Bernardo, H. Qi, J. M. Shultz, A. M. Cohen, N. F. Johnson, and P. S. Dodds. Social media affects the timing, location, and severity of school shootings. *arXiv preprint arXiv:1506.06305*, 2015.

[21] L. Mercken, T. A. B. Snijders, C. Steglich, E. Vartiainen, and H. de Vries. Dynamics of adolescent friendship networks and smoking behavior. *Social Networks*, 32(1):72–81, 2010.

[22] C. W. Mills. JSTOR, 1957.

[23] A. Smith and M. Garnier. *An Inquiry into the Nature and Causes of the Wealth of Nations*. W. Strahan and T. Cadell, London, 1776.

[24] P. Romer. Endogenous technological change. *J Pol Econ*, 98:S71–S10, 1990.

[25] G. M. Grossman and E. Helpman. Quality ladders in the theory of growth. *The Review of Economic Studies*, 58(1):43–61, 1991.

[26] R. Belderbos, S. Du, and D. Somers. Global Cities as Innovation Hubs: The Location of R&D investments by Multinational Firms. page 30, 2014.

[27] M. P. Allen. The Structure of Interorganizational Elite Cooptation: Interlocking Corporate Directorates. *American Sociological Review*, 39(3):393, 1974.

[28] `https://github.com/uvacorpnet`.