The organization, exchange and transfer of knowledge in socio-technical systems

Giacomo Vaccario[1]
[1]PhD candidate, Chair of Systems Design, ETH Zurich

1 Introduction

What is Knowledge?

We aim to investigate knowledge in the domain of science and economics. Tough an unambiguous definition of knowledge is disputed, we now discuss the meaning of this concept in both domains.

In economic theory knowledge would typically be considered as "human capital" (Walsh, 1935; Mincer, 1958; Schultz; Becker, 1964; Romer, 1994; Jaffe and Trajtenberg, 2002). This "capital" should not only be seen as a bare collection of goods, for it is often embedded in a social context that can affects strongly its properties and value (Coleman, 1988; Dolfsma, 2008; Scharnhorst *et al.*, 2012). To include this characteristic, some scholars consider knowledge as *"a dynamic framework connected to cognitive structures from which information can be stored, processed and understood"* (Howells, 2002; Scharnhorst *et al.*, 2012). By this, we explicitly state that knowledge involves cognitive structures, such as humans, that can be embedded in a social context and hence be affected from it.

The above concept of knowledge comes from economics (Howells, 2002) and it is also accepted by scholars modeling the structure and evolution of science (Scharnhorst *et al.*, 2012). For this reason, it well matches the general aim of this project, that is investigating knowledge in the domain of science and economics.

Data, information and knowledge

When defining *knowledge* we used the term *information*. These are two distinct, but connected concepts. A detailed discussion about their differences goes beyond the scope of the PhD project presented in this research plan. At the same time, let us clarify how we interpret and use these terms with an example.

Assume that Alice (A), a phd student, passes a list with 20 dates to Prof. Bob (B), her supervisor. This list of dates for B is just some *data*, i.e. it is a collection of values without a specific meaning.

If B asks to A:"Which are the publishing dates of my 20 most cited articles?" and A passes a list with 20 dates to B, B has received some *information*. Now, the list is not a bare collection of values, instead it has a meaning that depends on the question asked. To oversimplify, information is an answer to a question, while data is not.

Lastly, Prof. B wants to understand why his "top articles" were published during some specific years. To do this, for example, he searches for patterns in the age distribution of these articles. By finding and understanding these patterns, he would gain some *knowledge*. Hence Prof. B has to store, process and understand information in order to discover and acquire new knowledge.

The organization, exchange and transfer of knowledge.

There exists a vast literature studying the organization, exchange and diffusion of knowledge. Important contributions are Jaffe (1986); Nonaka (1994); Nonaka and Takeuchi (1995); Jaffe and Trajtenberg (2002); Antonelli (1996); Owen-Smith and Powell (2004); König *et al.* (2011); Tomasello *et al.* (2015) in the economic domain and Lotka (1926); Price (1951, 1976); Noyons and van Raan (1998); Boyack *et al.* (2005); Gargiulo *et al.* (2016); De Domenico *et al.* (2016) in the scientific domain.

In particular, Jaffe and Trajtenberg (2002); Tomasello *et al.* (2015) use patent data to quantify innovation and measure knowledge flows. Indeed, knowledge can be encoded in patents and by this patents become knowledge artifacts. In the attempt of establishing different types of knowledge, Polanyi (1966) distinguished between *explicit* and *tacit* knowledge. By following this point of view, we assume that only the explicit knowledge is encoded in knowledge artifacts (Polanyi, 1966; Nonaka, 1994; Nonaka and Takeuchi, 1995; Howells, 2002). With patents and their relations, we can investigate the organization and evolution of knowledge. The possible relations between patents are many and of various types. Obvious examples include the citations among patents. While less obvious ones are the geographical distances between the inventors filing the patents.

In the field of *scientometrics* and *information science*, scholars use an equivalent approach. They consider scientific publications as knowledge artifacts and by analyzing these they study knowledge in science. This approach has been used to drawn maps of the scientific knowledge (Shiffrin and Börner, 2004; Boyack *et al.*, 2005; Leydesdorff and Rafols, 2009) and to models its evolution (Price, 1976; Börner *et al.*, 2004; De Domenico *et al.*, 2016).

As knowledge can be extracted from information, authors often infer knowledge diffusion by looking at the spreading of information or of other abstract quantities (Scharnhorst *et al.*, 2012). Examples are the models for the spreading of ideas (Goffman and Newill, 1964), innovation (Rogers, 1962), technologies (Mahajan and Peterson, 1985), news and rumors (Bartholomew, 1967). At the same time humans produce and acquire knowledge, and by moving from one place to another, they spread it. For this reason, the physical migration of humans is an important form of knowledge diffusion studied in various works (Vlachỳ, 1981; Le Pair, 1980; Gargiulo *et al.*, 2016).

As in this PhD project we will focus of on the mobility of people to proxy the spreading of knowledge, we are tempted to use expressions such as *mobility of knowledge* or *migration of knowledge*. At the same time, these terms could be easily misinterpreted. The term "migration" usually refers to the movement of humans to a new country or to the seasonal movement of animals. Whereas the term "mobility" is mostly used in the field of physics to describe how quickly particles move in materials (e.g. electron mobility). In a similar sense, the term "mobility" is also used to discuss about human dynamics (e.g. urban mobility). To avoid such confusion, we define for this project *transfer of knowledge* as the spreading of knowledge occurring thanks to the physical migration of people. For example, scientists move and transfer knowledge from one university to another, or inventors move and transfer knowledge from one firm to another.

Academia and the R&D alliances as socio-technical systems.

In modern society there are many systems in which social and technical aspects are strongly intertwined. Prominent examples are academia, meaning scientists with their research activities, and R&D alliances, meaning partnerships arranged by firms to promote their R&D activities. In these systems, the social actors (scientists and firms) influence each other depending on both their social and technical characteristics. For example, in academia, scientists' technical skills and knowledge are important for their scientific impact. At the same time, scientists' scientific impact is correlated with their centrality in the social network reconstructed using their co-authorship activities (Sarigöl *et al.*, 2014). In economics, firms select R&D partners depending on their social and technological capitals (Burt, 1992; Podolny, 1993; Gulati, 1995; Rosenkopf and Nerkar, 2001). This interdependence between social and technical aspects makes academia and the R&D alliances two socio-technical systems.

The impact of knowledge.

Knowledge is considered one of the key production factors explaining wealth in modern capitalist societies (Walsh, 1935; Mincer, 1958; Schultz; Stewart and Ruckdeschel, 1998). The competitive advantage of firms depends more and more on the their investments in knowledge-based capital. In addition, an important share of the market value of many leading firms reflects their knowledge assets, such as patents and know-how (OECD, 2013). The importance of knowledge is not restricted to economics; for example, one of the main scope of Science is to organize and create knowledge. Obviously, the scientific progress has a direct impact on many aspects of people's lives. Hence assessing, categorizing and measuring knowledge have a broad relevance in our society. Building on the extant literature on this subject, we aim at developing new tools for quantifying knowledge and for studying its exchange and transfer.

2 Research Gap

To clearly define the research gap, we start by introducing two "macro-research questions" together with relevant examples that can be used to investigate them. Then we present three concerns that are discussed in the literature and are linked to the "macro-research questions"

The three concerns are related, but not restricted to limitations arising from network-analytic methods.

For this PhD project we concentrate on two specific classes of problems. The first class of problems is related to the question of *how knowledge artifacts are linked to each other, ranked and filtered in repositories*. Examples of these repositories are patent and scientific publication databases. The second class of problems relates to the question of *how knowledge is exchanged and transferred*. Indeed, knowledge is not only produced and encoded in knowledge artifacts, as patents or scientific publications, but it is also exchanged by human beings and it diffuses in our society. R&D alliances among firms and co-authorship of papers among scientists are examples of activities favoring knowledge exchange. While the physical migration of inventors or scientists is an example of knowledge transfer (see Introduction for more details).

Answering the questions highlighted in the above paragraph is a multidisciplinary effort made by many researchers. In particular economists, bibliometricians and computer scientists contributed by proposing tools for managing knowledge artifacts, modeling knowledge exchange and its transfer.[1] Successful tools have been developed by applying network-analytic methods (Pinski and Narin, 1976; Brin and Page, 1998; Tomasello *et al.*, 2015; Clauset *et al.*, 2015; Gargiulo *et al.*, 2016). The use of such methods was possible as the data could be represented using a network abstraction. For example, a scientific publication database can be represented as a *journal-journal citation network*, that is a network where a node is a scientific journal and a link between two journals is a citation between two papers published in these journals (Pinski and Narin, 1976). Similarly, data about scientists' careers[2] can be represented as *faculty hiring network*, that is a directed network where each node is a university and a directed link represents a scientist graduating from a (source) university and, becoming a faculty member in a target university (Clauset *et al.*, 2015)

Even though tools developed using network-analytic methods have been proven to be useful, recent advances in data mining and network theory raised concerns about their naive application to complex data (Butts, 2009; Zweig, 2011; Mariani *et al.*, 2015; Scholtes, 2017). In the following sections we will give an overview of three important concerns.

2.1 The application of network-analytic methods to citation data

One established method to extract knowledge from patent and scientific publication data is *citation analysis*. It is used to investigate citation patterns in order to disclose properties of documents. In citation analysis, we use network-analytic methods as we can represent documents and their citations as a *citation network*. This is a network where the nodes are documents and

[1]When writing knowledge transfer, we always refer to how scientists or inventors move across firms, institution, cities or countries, unless otherwise specified.

[2]Here, we exactly mean the personal route/trajectory of a scientist in his/her professional life.

the links are citations among documents. When representing citation data as a network, we discard many proprieties of the documents, such as their age or their topic of pertinence.

At the same time, the cumulative number of citations received by documents strongly depends on their age and topic of pertinence (Schubert and Braun, 1986; Vinkler, 1986). Then a question arises: *how can we use citation analysis to compare documents of different age and belonging to different topic?* Answering this question is important for many reasons. For example, we increasingly use citation analysis for research evaluation, but we cannot use it fairly to compare researches from to different topics.

Various works proposed an answer to the above question (see Waltman (2016) for a recent review). A simple, yet successful method is to divide each document's citation count by the mean number of citations for documents in the same topic published in the same year (Radicchi *et al.*, 2008). By this procedure we obtain a new indicator called *relative citation count*. Radicchi *et al.* (2008) claim it can be used as "*an unbiased indicator for citation performance across disciplines and years*" for scientific publications. Subsequent works challenged this finding (Albarrán *et al.*, 2011; Waltman *et al.*, 2012), which leaves the debate on age- and field- normalization procedures still open. In addition, to the best of our knowledge, a statistical test to assess if indicators are simultaneously age- and field-normalized is still missing.

2.2 Collaboration networks and knowledge: Which is the cause-and-effect relationship?

Researchers from many disciplines have being addressing the long lasting question of how we collaborate. A network model that shed new lights on this question in the economic domain was presented by Tomasello *et al.* (2014). In this work, the authors successfully use an agent-model to reconstruct the R&D network between firms and hence, their collaboration patterns. Building on such network model, Tomasello *et al.* (2015) investigate how firms exchange knowledge during collaborations. Interestingly, the authors find that firms exchange knowledge at low rate, meaning that knowledge is rather a determinant than a consequence of collaborations. This result is in line with the fact that firms use collaborations to access new knowledge (Nooteboom, 1999); but then firms do not use this knowledge to expand their knowledge base.

Tomasello *et al.* (2015) assign a knowledge position to firms by looking at their patents portfolios. In particular, the authors consider the main 8 sections of the International Patent Scheme (IPC) to classify patents. This forces the knowledge space to have 8 dimensions. Such choice raises two technical, but rather critical problems. *First*, the model of Tomasello *et al.* (2015) is a particular type of a n-vector model in 8 dimensions (Stanley, 1968). Then, its properties and behaviors are linked to its dimensionality. Hence the results obtained using the IPC might not hold when using another classification scheme with a different number of dimensions. *Second*, it is reasonable to doubt that the 8 dimensions of the IPC sections can adequately describe the knowledge space: i) IPC sections are broad (i.e., sections include diverse technologies that rather unlikely share the same knowledge base), ii) some technological fields are dispersed across several IPC sections (e.g. food chemistry related patents may be found in two different sections). To establish if the result of Tomasello *et al.* (2015) holds when using more refined patent classification schemes with higher dimension is still an open question.

In addition, the question of how social actors in other domains collaborate and exchange knowledge is unanswered. One could argue that the answer to such question changes with respect to the actors and the domain of activity, but there may be also evidence for common features across different domains. Indeed, a recent work shows that a unified modeling approach can reproduce and explain the structural and the dynamic features of collaborations in different domains (Tomasello *et al.*, 2017). Such work uses the same agent-based model of Tomasello *et al.* (2014) to investigate the collaboration patterns in economics and science. Precisely, the authors focus on collaborations aimed at the production of new knowledge: R&D alliances among firms from 6 different industrial sectors and co-authorship activities among scientists from 6 distinct disciplines of *physics*. Building on this result and on the methods presented in (Tomasello *et al.*, 2015), it is interesting to ask how much knowledge scientists exchange during co-authorship activities. In particular, it is interesting to check if the result of Tomasello *et al.* (2015), namely that knowledge is rather a determinant than a consequence of collaborations, holds in the scientific domain.

2.3 The loss of temporal correlations

The third concern arises as we often need to make the *transitivity assumption* in order to use network-analytic methods. By transitivity assumption we mean the following: when inferring (from data) the existence of links from a to b and from b to c, we automatically permit the existence of a path from a to c via b.[3] Recently, Scholtes *et al.* (2014) have shown that this assumption is not justified in many real-life systems as non-trivial temporal correlations of events invalidate it. For this reason, network-analytic methods can be inadequate to investigate time sequences of events.

Recent advances in data mining and network theory overcame some of the limitations caused by the transitivity assumption. In particular, in (Scholtes *et al.*, 2014, 2016; Scholtes, 2017) the authors show how to correctly construct networks and study diffusion processes by using data containing time sequences of events. In addition, Scholtes (2017) provides a statistical test for doing model selection that accounts for possible temporal correlations (present in the data). And in this test a network is one of the candidate models. Hence, we can now verify when it is justified to interpret data using a network abstraction and when it is not.

[3] From a mathematical point of view this assumption is equivalent to the following statement: after inferring from the data a set of nodes and a set of links to construct a network and its adjacency matrix, the n^{th}-power of the adjacency matrix is the set of paths of length n allowed on the network.

With the methods provided in (Scholtes *et al.*, 2014, 2016; Scholtes, 2017) we can investigate two sets of problems: *old* problems that might have been tackled incorrectly by using a network perspective, i.e. relevant temporal correlations might have been discarded, and *new* problems for which temporal ordering is important, but have not been tackled yet. In both these sets, we are interested to those problems related to knowledge and its transfer.

3 Research Questions

We now provide 7 research questions (RQs) whose answers will permit to fill the above stated research gap.

3.1 The organization of Knowledge

RQ1: Assessing multiple normalizations. We need to age- and field- normalize citation-based indicators in order to compare documents of different age and from different fields. How can we assess that these indicators have been *simultaneously* age- and field- normalized?

To answer this question we propose to develop a statistical tests to quantify ranking biases.

RQ2: Developing a new normalization procedure. As the procedure of Radicchi *et al.* (2008) failed to correctly age- and field-normalize citation count and to the best of our knowledge there are no better ones, how can we develop a better one?

To answer this question we propose a re-scaling similar to the one of Mariani *et al.* (2016). In this work, the authors age-normalize citation-based indicators using a procedure based on the z-score. We aim to extend such procedure to simultaneously age- *and* field-normalize indicators.

RQ3: developing new knowledge order. How can we use time-correlations present in citation data to develop citation-based indicators?

To answer this question we aim to use the approach of Scholtes (2017) to develop new citation-based indicators. In particular, we will develop indicators at the journal-level, i.e. indicators to rank journals. Journals have different focuses representing distinct field of science. Hence, by considering time-correlations in the citation patterns at journal-level, we investigate which scientific field fosters the research of other fields.

3.2 The exchange of knowledge

RQ4: knowledge exchange among firms. The results of Tomasello *et al.* (2015) indicate that knowledge is rather a determinant than a consequence of R&D collaborations. How does this result change when using different methods to quantify knowledge?

To answer this question we will use different patent classification schemes to embed the firms in new knowledge spaces. These spaces will have also different number of dimensions. By this we will investigate if the results of Tomasello *et al.* (2015) are independent of the dimensionality of the knowledge space and of the classification scheme used.

RQ5: knowledge exchange among scientists. How can we extend the model and the analysis of Tomasello *et al.* (2015) to the scientific domain?

The model presented by Tomasello *et al.* (2015) relies on an agent-based model made of two parts: first it reconstructs the collaboration network among firms and then it simulates the knowledge exchange. Hence, to answer RQ5, we have to start by reconstructing the collaboration network between scientists. Then, we will simulate the knowledge exchange among scientists. To do this, we will embed the scientists in a knowledge space by using their publication lists. The knowledge space will be defined by the classification scheme used to categorize the publications.

As final remark, it is worth to noticing that it is possible to reconstruct the collaboration network between scientists by using data about their co-authorship activities (Tomasello *et al.*, 2017). By using this approach, our investigation will have one main difference compared to (Tomasello *et al.*, 2015). In (Tomasello *et al.*, 2015), the author employs two distinct types of data to infer the collaborations among firms (R&D alliances data) and their knowledge positions (patents portfolios data). While we will use co-authored publications to infer both the collaborations among the scientists and their knowledge positions. To be precise, to calculate the knowledge positions we will use also single-author publications. For example in the APS dataset these are less than 5% of the total number of the listed publications. Clearly single-author publications cannot be used to infer collaborations. Hence, the data we will use to infer collaborations among scientists and their knowledge positions are slightly different. Possible limitations of this approach will be investigated. At the same time, we expect the approach to be reliable as it has been employed to investigate knowledge exchange among inventors (Singh, 2005).

3.3 The transfer of knowledge

Knowledge diffuses not only thanks to knowledge artifacts, such as patents or scientific publications, but also thanks to physical migrations of people. In particular, in academia scientists produce and acquire knowledge which can be encoded in scientific publications. At the same time, they also move from one research institute to another and hence, spread their knowledge across different places. For this reason we now propose two question on academic mobility in order to investigate the transfer of knowledge.

RQ6: temporal correlations in the transfer of knowledge. How should we model scientists' academic mobility in order to retain temporal correlations and a network perspective?

Many methods and models rely on the assumption that knowledge diffuses on interconnected structures, such as networks. Thanks to the recent results of Scholtes (2017), we can question this assumption and ask if it is confirmed or invalidated by data. Clearly, the answer to this question depends on the knowledge diffusion process investigated. For this reason, we focus on one type of knowledge diffusion: the mobility of scientists. Precisely, we will investigate how scientists in academia move from one research institutes to another. This investigation will be performed using the methodology of Scholtes (2017). By this, we will first understand if it is correct to assume that knowledge migrates on an underlying faculty hiring network (Clauset *et al.*, 2015). Secondly, we will investigate how temporal correlations change university ranking based on scientists' mobility data (Clauset *et al.*, 2015).

RQ7: new agent-based model for knowledge transfer. Let us assume that temporal correlations in the migration trajectories of scientists break the transitivity assumption. Then, how can we use an agent-based model to reproduce these type of trajectories?

We aim to answer to this question in two steps. First we will find the minimal ingredients (i.e., assumptions, microscopic rules, and parameters) needed for an agent-based model to reproduce trajectories that should be interpreted using higher-order networks. Secondly, we will map these ingredients to aspects of real-world migration processes. By this, we will be able to calibrate and validate our ABM using data on academic mobility. We give more details about this ABM in Sect. 5.

4 Progress

RQ1. Done. The results have been published in a peer-reviewed journal (Vaccario *et al.*, 2017a).

RQ2. Done. The results have been published in a peer-reviewed journal (Vaccario *et al.*, 2017a).

RQ3. We have completed the literature research on measures and algorithms defined on data with temporal correlations. In addition, we have tested on synthetic data a first set of measures. More refined measures still need to be developed and applied to real-world data.

RQ4. Done. The results have been submitted in a peer-reviewed journal (Vaccario *et al.*, 2017b). In addition, they will be presented to the upcoming conference Kreyon 2017.

RQ5. The agent-based model presented in (Tomasello *et al.*, 2015) is available and also the necessary co-authorship data. We still have to decide which classification schemes to use to classify scientific publications and embed their authors in a knowledge space.

RQ6. We already collected and processed the data. The result shows that a second-order network is needed to explain the migration of scientists inside the top-100 universities. The top-100 universities have been decided using the ranking given by *Times Higher Education*. Other rankings can be used and no strong difference are expected as the top-100 positions in university rankings are usually occupied by the same set of universities. We aim to present such result in a peer-reviewed journal.

RQ7. The key ideas of the model have been discussed orally. However, the proposed ABM and its application still have to be developed and tested using real-world data. In the next section, we give some details about this ABM.

5 An ABM for academic mobility

5.1 How many types of agents do we need?

For answering RQ7, we have to reproduce trajectories on a discrete space. These trajectories will represent scientists moving and the discrete space will represent the research institutes. One could argue that the only agents active in this process are the scientists as they move from one research institute to another. At the same time, their movement is conditional upon the acceptance of research institutes. Hence also research institutes have an active role. Before defining the ABM, we need to first clarify how many types of agents we need. The possible choices are: one type of agents, either scientists or research institutes, or two types of agents, research institutes and scientists.

We propose to have only one type of agents (scientists) that move in a discrete space (defined by the research institutes). This is the best choice as it will give results that we can interpret more clearly. Indeed, to analyze an ABM with one type of agents is obviously more simple than to analyze one with two types. Also, we find more intuitive to have the scientists as active agents. Recall that our final goal is to investigate the transfer of knowledge. By assuming the scientists as the (active) agents for the ABM, they have an active role in producing and transferring their knowledge in the system. Whereas, if we take the research institutes as (active) agents, we have the scientists actively producing knowledge, but then they are passively exchanged by the research institutes. This scenario becomes more complicated as we have to understand the interplay between scientists, research institutes and knowledge together with their passive and active relations. Hence, we propose the scientists as the only (active) agents in our ABM.

5.2 Life cycles

We argue that a scientist just starting his/her career is different from a scientist that has been working for 20 or more years. For example, we expect experienced professors with a tenure position to change their affiliations less frequently than younger scientists, such PhD students or PostDocs. We also argue that a good proxy for determining the career stage of a scientist is the time elapsed from its first publication. We name this time the "academic age" of a scientist. This will be an attribute of our agents that will be used to characterize them with their behavior. In addition, we had some preliminary results that allow to use the "academic age" of scientists to classify them in three categories: PhD, PostDoc, Professors. This motivates us to divide our agents in three categories.

Recently, we have identified two distinct data sets (*MAG* and *ORCID data*, see Sect. 7) that permit to verify *i*) if scientists mobility depends on their "academic age" and *ii*) how stable our classification is. Hence, we propose to do the above two checks. In case we have a positive result, we will include "academic age" as internal degree of freedom of the agents and we will divide agents in different categories reflecting our classification depending on their "academic age".

As last remark, note that science is an open system meaning that scientists move in and out from it. Indeed, graduate students and people in general start research careers at any time of all years. Whereas, the reasons driving a scientist outside this system are various, but we can divided them in two: either he/she finds a work position outside science (e.g., in a private firms) or he/she stops working (e.g., he/she goes to pension). To replicate this characteristic, our agents will become active and explore research institutes at different time during the simulations. And, at a certain time they will become inactive and leave the system.

5.3 The utility function

Our ABM will be developed using *reflective* agents. This means that agents will be assigned with an utility function that describes how the they evaluate alternative options. In our case, this utility function permits to simulate scientists' decision process when evaluating *if* to move to a new research institute. Hence, to define the utility function we now need to discuss: i) scientists' motivations driving their decisions and ii) which are the constraints on their movements.

Scientists are driven by different motivations when doing science (Hagstrom, 1965; Gustin, 1973; Merton and Storer, 1973; Van Raan, 2000). In the literature, the motivations driving humans to engage in activities have been divided between *internal* (also called intrinsic) and *external* (also called extrinsic or instrumental)(Ryan and Deci, 2000; Smith *et al.*, 2014; Wrzesniewski *et al.*, 2014). We argue that both these types of motivations drive also scientists and their mobility. Indeed, on one hand a scientist changes his/her working place to do science in this place and hence, we expect that motivations internal to science play a role. On the other hand, changing research institution often means to change city, country or even continent. Clearly such changes affect the life of a scientist outside its work and hence, also motivations external to science play a role. We will capture both these motivations driving scientists inside variables which will define their utility function. We will also try to separate the importance of *internal* and *external* motivations by using distinct variables. This would permit to quantify which types of motivations are more important.

If on one hand it is true that scientists move from one research institute to another, on the other hand we all know that moving, and in general changes, imply costs. For example when moving to a new city, it is necessary to spend time and money to find a new apartment and to organize the travel. Moreover, once there it is necessary to reestablish new social relations inside and outside the working place. If the scientist has a family or a partner, then the costs increase as also the family or the partner have costs related to this change (e.g., finding a new school for the children and/or a new job for the partner). For this reason, we assume that the scientists are adverse to changes as any other humans. To capture this in the agents' utility function, we impose that an agent moves to a new institute when her expected gain from the movement is higher than the costs.

To conclude, we discuss an important constraint on the mobility of scientists. A scientist can apply and move to a new research institute when a position gets opened. Or, thanks to specific grants he/she can propose him/her-self as researcher in the institute. However, even in this case the institution still has to accept the scientist as future employee. From the research institute perspective, this is motivated by the fact that there are costs to interact/interface with a scientist even if he/she comes with its own funding. For this reason, we assume that the motions of scientists are triggered and permitted thanks to openings of job positions occurring in the research institutes. This assumption is reflected in our model as a constraint in the mobility of the agents. During the simulations they will be able to move only to those nodes that have open positions.

5.4 Features of the utility function

We now describe the features that we propose to include in the agents' utility function and how to measure them from data.

There are many reasons why a scientist chooses an university over another and we distinguish between motivations external and internal to science (see Sect. 5.3). We start with the former which can be identified by taking an economic perspective. We argue that a scientist prefers research institutes where productivity, research quality and salaries are higher. We propose to proxy productivity and research quality respectively with the number of publications from institutes and the number of citations these articles receive. From our preliminary analysis, we find that the number of publications from an institute are correlated with the number of researchers working in the institute, i.e. its size. Likewise, we find that also the total number of citations received by publications from a research institute is correlated to its size. For this reason, we will re-scale these two quantities by the size of the research institutes. In addition, by answering RQ1 and RQ2, we find that the average number of publications and citations of a scientist depend on the scientific field were he/she works. Therefore, we will have to also account for this.

The salaries payed to scientists vary a lot across American, European and Asian research institutes and are paid in different currencies. Moreover, the raw value of a salary (lets say in dollars) can be a misleading feature as the purchase power of 1$ varies from city to city. We argue that these variations reflect the life standards where the research institutes are located. Hence, instead of considering the raw numbers quantifying the salaries, we will rank the city in which institutes are located and assign to each institute the rank of its city. By this, if an institute is in a city with a higher rank, we assume that it pays higher salary after accounting for differences in living costs. Whereas, if two institutes are in the same city, we will assume that they will pay the same salary. A ranking of cities that we are considering to use is the *Mercer Quality of Living Ranking*.[4]

As mentioned in Sect. 5.3, a scientist moves from a (source) research institute to another (target) one not only for socio-economical reasons. He, or she, is driven also by motivations internal to science. To proxy these motivations we propose the following features: the pre-existance of collaborations between the source and target institutes and personal collaborations with scientists working in the target institute. We argue that these are features internal to science as both types of collaborations represent aligned research interests and similar working values. We propose to infer the collaborations links between institutes and scientists by using data on co-authorship activities.

As final remark, let us propose how to test which of the above features are relevant and how to estimate their relative importance. There are three different approach that we rank from more traditional to less traditional: i) survey, ii) stochastic actor-based modeling and iii) data mining. All these approaches have their advantages and shortcomings. Following the competences of the *Chair of System Design*, we will focus on the last two.

5.5 Validation of the model

We propose an ABM to reproduce the trajectories of scientists. In RQ7, we ask that this model should capture especially one aspect of these trajectories: the correct temporal correlations. In our preliminary results we find that scientists' trajectories across top-100 research institutes should not be interpreted using an aggregated network model. Instead, the correct model to use is a second-order network model, i.e. a higher-order network of order 2. This indicates the presence of non-trivial temporal correlations in the trajectories (Scholtes *et al.*, 2014; Scholtes, 2017). Hence, we will validate our ABM against this property observed in the data. Precisely, our model has to reproduce scientists' trajectories across research institutes and the correct model to interpret these trajectories has to be a second-order network model.

[4] `https://mobilityexchange.mercer.com/Portals/0/Content/Rankings/rankings/qol2017e784512/index.html`

6 Linking back the research questions to the overarching perspective

Knowledge is embedded in a social context

We contribute to the study of the organization of knowledge by analyzing how knowledge artifacts are linked to each other, ranked and filtered in repositories. Precisely, we analyze a new scholarly publication database, the *MAG* (see Sect. 7 for details), and test a set of citation-based indicators. Our contribution goes beyond the simple analysis of a database as we introduce a novel statistical test and a general normalization procedure. These permit to respectively quantify and suppress ranking biases coming from the age of the papers and their field of belonging. By this, we have developed new analytic methods that quantify and suppress ranking biases coming from the social context in which the knowledge artifacts were produced. Our new methods have been presented in (Vaccario *et al.*, 2017a).

The use and importance of developed methods are clear when recalling our discussion about the concept of knowledge in the economic and scientific domain (See 1). There we emphasize that knowledge is embedded in a social context. By using our new methods, we can quantify to which extent this characteristic of knowledge influences rankings of papers based on citations. For example, we find that the distributions of citation counts of papers with different age and belonging to different fields are far from having a universal functional as claimed by Radicchi *et al.* (2008). This points out that the mechanisms motivating scientists to cite other papers vary over time and across fields. In addition, with the convention that papers encode knowledge, we argue that the citation differences reflect the different social contexts in which the knowledge (behind the cited and the citing papers) is embedded.

In Sect. 1 we discuss the concept of knowledge in the economic and scientific domain. There we emphasize that knowledge is embedded in a social context. In (Vaccario *et al.*, 2017a), our results reflect this property of knowledge. We find that the distribution of paper citation counts is different depending on the scientific field and on the age of the papers. With the convention that papers encode knowledge, we argue that the citation differences reflect the different social contexts in which the knowledge (behind the cited and the citing papers) is embedded.

In particular, we find statistically significant differences in the tails of the distributions of paper citation counts from different scientific field and of different age. Usually the tails of distributions are not of particular interest as they represent rare events. Whereas in scientometrics and citation analysis they are more interesting. They represent sets of top-cited papers, i.e. papers that have attracted a lot of interests from the scientific community. Statistically significant variations in these tails across scientific fields and years point out that the breath of scientists' interests changes over time and across fields of knowledge. Following the perspective of Van Raan (2000), we argue that the motivations triggering these changes are external to science and are stimulated by socioeconomic problems. However, we leave the test of this hypothesis to future research.

Quantifying knowledge exchange

As the properties of knowledge depend on the social context in which it is embedded, we have proposed to investigate the "macro-research question" of *how knowledge is exchanged* in different systems (see Sect. 2 and 3.2). In particular, we focus on socio-technical systems and consider R&D alliances and academia as representative examples (see Sect. 1). To answer the above "macro-research question" in these systems, we used the approach of Tomasello *et al.* (2015). Precisely, we have first confirmed its main result, i.e. firms exchange knowledge at extremely low rate during R&D alliances (Vaccario *et al.*, 2017b). Then, we aim to check if this result holds also in academia. This part of the PhD project still have to be tackled.

To confirm the main result of Tomasello *et al.* (2015), we used the ISI-OST-INPI patent classification scheme to define a 35 dimensional knowledge space. In this space, we have embedded firms based on their patent portfolios and computed the knowledge distances among allied firms. Distances between firms before and after R&D alliances are respectively called pre- and post-alliances distances. Then, we modified the ABM of Tomasello *et al.* (2015) to simulate the knowledge exchange process occurring during collaborations. We used the pre-alliance distances as input of our model, while the post-alliance distances to calibrate it. From the calibration procedure, we find that firms exchange knowledge extremely slowly during R&D alliances. I.e., our ABM simulations better match the observed post-alliance distances when the parameter modeling knowledge exchange frequency is small. This result of our ABM simulations is qualitatively identical to the one of Tomasello *et al.* (2015).

From an economic perspective, this result suggests that firms' knowledge is not a bare collection of goods which can be easily moved or multiplied from one firm to another.On the contrary, a firm has some specific knowledge that acts as an almost unchangeable base. Hence, knowledge resides at the core of a firm. We argue that this result supports the choice of including knowledge when estimating the market value of a firm.

Also, in (Vaccario *et al.*, 2017b) we have investigated if an *optimal knowledge distance* that fosters the R&D alliances exists (Nooteboom *et al.*, 2007). We find that the distribution of pre-alliance distances significantly differs from the distribution of knowledge distances among all possible pairs of firms. Moreover, we find that the mode of the former distribution (the knowledge distance measured more often before an alliance was observed) was significantly different from zero and smaller than the mode of the latter distribution. This means that there is an observable and significant difference in the knowledge base of firms engaging in alliances. But this difference is smaller compared to what it is expect if firms were forming alliances without accounting for the knowledge of their future partners. This result points in a direction: the knowledge distance among firms is definitely a determinant of collaborations. At the same time, there is no evidence of the existence of one optimal knowledge distance. It rather exists a *set of distances* that foster R&D alliances.

The transfer of knowledge

As final remark let us comment regarding our investigation of knowledge transfer that we proxy by the mobility of scientists. On this topic, we only have preliminary results. We find that trajectories of scientists are best represented by higher-order networks (Scholtes *et al.*, 2014). In other words, the institution where a scientist will go depends not only on the current institution where he/she works, but also on the previous ones where he/she was. This result challenges the idea that one scientist when leaving one institution he/she has access to any other ones. On a macroscopic level, this means that the transfer of knowledge, proxied by the mobility of scientists, does not occur on an underlying *fully* connected network of research institutes. Instead, the scientists together with their knowledge follow only specific paths. We still have to investigate the motivations and the implications of such result.

7 Data sets

For the research project presented in the previous sections data will have high relevance. Below we provide a list of data sets which may be used in this project. The list is not necessarily complete. However, it contains the dataset used for exploratory research purposes and to answer RQ1, RQ2 and RQ4.

Microsoft Academic Graph (*MAG*). Scholarly publication database that contains more than 126 millions of publications and more than 467 millions citations. Each publication is also endowed with various properties such as unique ID, publication date, title, journal ID, etc. For more details about this data set see (Sinha *et al.*, 2015; Harzing and Alakangas, 2016; Hug and Braendle, 2017; Vaccario *et al.*, 2017a))

ORCID data. ORCID is a noprofit that provides services to researchers and research institutes. Recently, Bohannon and Doran (2017) has used information provided by ORCID to create a database containing profiles of researchers, i.e. their CVs. The database is available at `http://dx.doi.org/10.5061/dryad.48s16`.

APS Data. It provides a list of papers published in any American Physical Society (APS) journals, namely Physical Review Letters, Reviews of Modern Physics, and all Physical Review journals. Together with the publications it provides the PACS codes of the papers (a classification scheme) and publications date. This data is freely available for research at `https://journals.aps.org/datasets`.

NBER data. Patent Citations Data of the U.S.A. National Bureau of Economic Research (NBER). It contains detailed information on patents granted in the U.S.A. and other contracting countries, from 1971 to present.

SDC Platinum database. This contains data about 672,000 announced alliances from all countries between 1984 and 2009 with daily resolution. The economic actors participating in these alliances are of several types, e.g. investors, manufacturing firms and universities. Each actor listed in the data set is associated with a SIC (Standard Industrial Classification) code that allows us to unambiguously assign its corresponding industrial sector. This database was obtained from *Thomson Reuters*(`http://thomsonreuters.com/sdc-platinum/`).

References

Albarrán, P.; Crespo, J. A.; Ortuño, I.; Ruiz-Castillo, J. (2011). The skewness of science in 219 sub-fields and a number of aggregates. *Scientometrics* **88(2)**, 385–397.

Antonelli, C. (1996). Localized knowledge percolation processes and information networks. *Journal of Evolutionary Economics* **6(3)**, 281–295. ISSN 1432-1386. URL `http://dx.doi.org/10.1007/BF01193634`.

Bartholomew, D. J. (1967). *Stochastic models for social processes.* Wiley New York.

Becker, G. (1964). *Human capital: a theoretical and empirical analysis, with special reference to education.* National bureau of economic research publications: General series, National Bureau of Economic Research; distributed by Columbia University Press.

Bohannon, J.; Doran, K. (2017). Introducing ORCID. *Science* **356(6339)**, 691–692. ISSN 0036-8075. URL `http://science.sciencemag.org/content/356/6339/691`.

Börner, K.; Maru, J. T.; Goldstone, R. L. (2004). The simultaneous evolution of author and paper networks. *Proceedings of the National Academy of Sciences* **101(suppl 1)**, 5266–5273. URL `http://www.pnas.org/content/101/suppl_1/5266.abstract`.

Boyack, K. W.; Klavans, R.; Börner, K. (2005). Mapping the backbone of science. *Scientometrics* **64(3)**, 351–374.

Brin, S.; Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems* **30(1-7)**, 107–117.

Burt, R. (1992). *Structural Holes: The Social Structure of Competition.* Harvard University Press.

Butts, C. T. (2009). Revisiting the foundations of network analysis. *science* **325(5939)**, 414–416.

Clauset, A.; Arbesman, S.; Larremore, D. B. (2015). Systematic inequality and hierarchy in faculty hiring networks. *Science advances* **1(1)**, e1400005.

Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology* **94**, S95–S120. ISSN 00029602, 15375390. URL `http://www.jstor.org/stable/2780243`.

De Domenico, M.; Omodei, E.; Arenas, A. (2016). Quantifying the diaspora of knowledge in the last century. *Applied Network Science* **1(1)**, 15. ISSN 2364-8228. URL `https://doi.org/10.1007/s41109-016-0017-9`.

Dolfsma, W. (2008). *Knowledge Economies: Organization, Location and Innovation.* Routledge Studies in Global Competition, Taylor & Francis. ISBN 9781134116669. URL `https://books.google.de/books?id=b4h748Rgh_cC`.

Gargiulo, F.; Caen, A.; Lambiotte, R.; Carletti, T. (2016). The classical origin of modern mathematics. *EPJ Data Science* **5(1)**, 26. ISSN 2193-1127. URL `https://doi.org/10.1140/epjds/s13688-016-0088-y`.

Goffman, W.; Newill, V. (1964). Generalization of epidemic theory. *Nature* **204(4955)**, 225–228.

Gulati, R. (1995). Social structure and alliance formation patterns: A longitudinal analysis. *Administrative science quarterly* **40(4)**, 619–652.

Gustin, B. H. (1973). Charisma, Recognition, and the Motivation of Scientists. *American Journal of Sociology* **78(5)**, 1119–1134. ISSN 00029602, 15375390. URL `http://www.jstor.org/stable/2776628`.

Hagstrom, W. O. (1965). *The scientific community.* New York: Basic Books.

Harzing, A.-W.; Alakangas, S. (2016). Microsoft Academic: is the phoenix getting wings? *Scientometrics* , 1–13.

Howells, J. R. L. (2002). Tacit Knowledge, Innovation and Economic Geography. *Urban Studies* **39(5-6)**, 871–884. URL `http://dx.doi.org/10.1080/00420980220128354`.

Hug, S. E.; Braendle, M. P. (2017). The coverage of Microsoft Academic: Analyzing the publication output of a university. *arXiv preprint arXiv:1703.05539* .

Jaffe, A. B. (1986). Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value. *The American Economic Review* **76(5)**, 984–1001. ISSN 00028282. URL `http://www.jstor.org/stable/1816464`.

Jaffe, A. B.; Trajtenberg, M. (2002). *Patents, citations, and innovations: A window on the knowledge economy.* MIT press.

König, M. D.; Battiston, S.; Napoletano, M.; Schweitzer, F. (2011). Recombinant knowledge and the evolution of innovation networks. *Journal of Economic Behavior & Organization* **79(3)**, 145–164.

Le Pair, C. (1980). Switching between academic disciplines in universities in the Netherlands. *Scientometrics* **2(3)**, 177–191.

Leydesdorff, L.; Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the Association for Information Science and Technology* **60(2)**, 348–362.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* **16(12)**, 317–323. ISSN 00430439. URL `http://www.jstor.org/stable/24529203`.

Mahajan, V.; Peterson, R. A. (1985). *Models for Innovation Diffusion. Quantitative Applications in the Social Sciences*, vol. 48. SAGE Publications, Inc.

Mariani, M. S.; Medo, M.; Zhang, Y.-C. (2015). Ranking nodes in growing networks: When PageRank fails. *Scientific reports* **5**.

Mariani, M. S.; Medo, M.; Zhang, Y.-C. (2016). Identification of milestone papers through time-balanced network centrality. *Journal of Informetrics* **10(4)**, 1207–1223.

Merton, R.; Storer, N. (1973). *The Sociology of Science: Theoretical and Empirical Investigations.* Phoenix books, University of Chicago Press. ISBN 9780226520926. URL `https://books.google.ch/books?id=zPvcHuUMEMwC`.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy* **66(4)**, 281–302. ISSN 00223808, 1537534X. URL `http://www.jstor.org/stable/1827422`.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization science* **5(1)**, 14–37.

Nonaka, I.; Takeuchi, H. (1995). *The Knowledge-creating Company: How Japanese Companies Create the Dynamics of Innovation.* Oxford University Press. ISBN 9780195092691. URL `https://books.google.ch/books?id=B-qxrPaU1-MC`.

Nooteboom, B. (1999). *Inter-firm Alliances: Analysis and Design.* Routledge. ISBN 9780415181549. URL `https://books.google.ch/books?id=-yP8IFFQm-oC`.

Nooteboom, B.; Van Haverbeke, W.; Duysters, G.; Gilsing, V.; Van den Oord, A. (2007). Optimal cognitive distance and absorptive capacity. *Research policy* **36(7)**, 1016–1034.

Noyons, E. C.; van Raan, A. F. (1998). Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research. *Journal of the American society for information science* **49(1)**, 68–81.

OECD (2013). *Supporting Investment in Knowledge Capital, Growth and Innovation.* OECD Publishing. URL `/content/book/9789264193307-en`.

Owen-Smith, J.; Powell, W. W. (2004). Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community. *Organization science* **15(1)**, 5–21.

Pinski, G.; Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Information Processing & Management* **12(5)**, 297–312.

Podolny, J. M. (1993). A status-based model of market competition. *American journal of sociology* **98(4)**, 829–872.

Polanyi, M. (1966). *The tacit dimension.* University of Chicago Press.

Price, D. d. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science* **27(5)**, 292–306.

Price, D. J. d. S. (1951). Quantitative measures of the development of science. *Archive Internationale d'Histoire de Sciences* **4(14)**, 86–93. URL `http://garfield.library.upenn.edu/price/pricequantitativemeasures1951.pdf`.

Radicchi, F.; Fortunato, S.; Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences* **105(45)**, 17268–17272.

Rogers, E. (1962). *Diffusion of innovations.* Free Press of Glencoe. URL `https://books.google.ch/books?id=zw0-AAAAIAAJ`.

Romer, P. M. (1994). The Origins of Endogenous Growth. *The Journal of Economic Perspectives* **8(1)**, 3–22. ISSN 08953309. URL `http://www.jstor.org/stable/2138148`.

Rosenkopf, L.; Nerkar, A. (2001). Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal* **22(4)**, 287–306.

Ryan, R. M.; Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology* **25(1)**, 54–67.

Sarigöl, E.; Pfitzner, R.; Scholtes, I.; Garas, A.; Schweitzer, F. (2014). Predicting scientific success based on coauthorship networks. *EPJ Data Science* **3(1)**, 1.

Scharnhorst, A.; Börner, K.; van den Besselaar, P. (2012). *Models of Science Dynamics: Encounters Between Complexity Theory and Information Sciences*. Understanding Complex Systems, Springer Berlin Heidelberg. ISBN 9783642230677. URL https://books.google.de/books?id=yrjt2l1xwkgC.

Scholtes, I. (2017). When is a Network a Network? Multi-Order Graphical Model Selection in Pathways and Temporal Networks.

Scholtes, I.; Wider, N.; Garas, A. (2016). Higher-Order Aggregate Networks in the Analysis of Temporal Networks: Path structures and centralities. *European Physical Journal B* **89(3)**, 1–15. ISSN 1434-6028. URL http://link.springer.com/article/10.1140%2Fepjb%2Fe2016-60663-0.

Scholtes, I.; Wider, N.; Pfitzner, R.; Garas, A.; Tessone, C. J.; Schweitzer, F. (2014). Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. *Nature communications* **5**, 5024.

Schubert, A.; Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics* **9(5-6)**, 281–291.

Schultz, T. W. (). Investment in Human Capital. *The American Economic Review* **51(1)**, 1–17. ISSN 00028282. URL http://www.jstor.org/stable/1818907.

Shiffrin, R. M.; Börner, K. (2004). Mapping knowledge domains.

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management Science* **(51)**, 756–770.

Sinha, A.; Shen, Z.; Song, Y.; Ma, H.; Eide, D.; Hsu, B.-j. P.; Wang, K. (2015). An overview of microsoft academic service (mas) and applications. In: *Proceedings of the 24th International Conference on World Wide Web*. ACM, pp. 243–246.

Smith, E. R.; Mackie, D. M.; Claypool, H. M. (2014). *Social psychology*. Psychology Press.

Stanley, H. E. (1968). Dependence of critical properties on dimensionality of spins. *Physical Review Letters* **20(12)**, 589.

Stewart, T.; Ruckdeschel, C. (1998). Intellectual capital: The new wealth of organizations.

Tomasello, M. V.; Perra, N.; Tessone, C. J.; Karsai, M.; Schweitzer, F. (2014). The Role of Endogenous and Exogenous Mechanisms in the Formation of R&D Networks. *Scientific Reports* **4**, 5679. URL http://www.nature.com/srep/2014/140714/srep05679/full/srep05679.html.

Tomasello, M. V.; Tessone, C. J.; Schweitzer, F. (2015). The effect of R&D collaborations on firms' technological positions. ISBN 978-88-96687-07-9. ISSN 2280-787X. URL http://www.ifkad.org/Proceedings/2015/papers/261_IFKAD2015.pdf.

Tomasello, M. V.; Vaccario, G.; Schweitzer, F. (2017). Data-driven modeling of collaboration networks: A cross-domain analysis. *arXiv:1704.01342* URL `https://arxiv.org/abs/1704.01342`.

Vaccario, G.; Medo, M.; Wider, N.; Mariani, M. S. (2017a). Quantifying and suppressing ranking bias in a large citation network. *Journal of Informetrics* URL `http://www.sciencedirect.com/science/article/pii/S1751157717300974`.

Vaccario, G.; Tomasello, M. V.; Tessone, C. J.; Schweitzer, F. (2017b). Quantifying knowledge exchange in R&D networks: a data-driven model. *Journal of Evolutionary Economics (submitted)* .

Van Raan, A. F. (2000). The interdisciplinary nature of science: theoretical framework and bibliometric-empirical approach. *Practising interdisciplinarity* , 66–78.

Vinkler, P. (1986). Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics* **10(3-4)**, 157–177.

Vlachý, J. (1981). Mobility in physics.

Walsh, J. R. (1935). Capital Concept Applied to Man. *The Quarterly Journal of Economics* **49(2)**, 255. URL `+http://dx.doi.org/10.2307/1884067`.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics* **10(2)**, 365–391.

Waltman, L.; van Eck, N. J.; van Raan, A. F. (2012). Universality of citation distributions revisited. *Journal of the American Society for Information Science and Technology* **63(1)**, 72–77.

Wrzesniewski, A.; Schwartz, B.; Cong, X.; Kane, M.; Omar, A.; Kolditz, T. (2014). Multiple types of motives don't multiply the motivation of West Point cadets. *Proceedings of the National Academy of Sciences* **111(30)**, 10990–10995. URL `http://www.pnas.org/content/111/30/10990.abstract`.

Zweig, K. A. (2011). Good versus optimal: Why network analytic methods need more systematic evaluation. *Central European Journal of Computer Science* **1(1)**, 137–153. ISSN 2081-9935. URL `https://doi.org/10.2478/s13537-011-0009-x`.

8 Authentication by Signature

The following persons appending their signature therewith verify the authenticity of this document and agree to its content.

Date	Prof. Dr. Dr. Frank Schweitzer, Referee
Date	Giacomo Vaccario, PhD candidate