Paper to be presented at the

DRUID Society Conference 2014, CBS, Copenhagen, June 16-18

Big Egos in Big Science: Unlocking Peer and Status Effects in the evolution of collaborative networks

Jacob Jeppesen
Copenhagen Business School
Department of Innovation and Organizational Economics
jj.ino@cbs.dk

Kristina Vaarst Andersen
Copenhagen Business School
Innovation and Organizational Economics
kva.ino@cbs.dk

Giancarlo Lauto
University of Udine
Dipartimento di Scienze Economiche e Statistiche
gl.ino@cbs.dk

Finn Valentin
Coepnhagen Business School
Department of Innovation and Organizational Economics
fv.ino@cbs.dk

Abstract

In this paper we investigate the micro-mechanisms governing structural evolution and performance of scientific collaboration. Scientific discovery tends not to be lead by so called lone ?stars?, or big egos, but instead by collaboration among groups of researchers, from a multitude of institutions and locations, having a diverse knowledge set and capable of tackling more and more complex problems. This prose the question if Big Egos continues to dominate in this rising paradigm of big science. Using a dataset consisting of full bibliometric coverage from a Large Scale Research Facility, we utilize a stochastic actor oriented model (SAOM) to analyze both network endogeneous mechanisms and individual agency driving the collaboration network and further if being a Big Ego in Big Science translates to increasing performance. Our findings suggest that the selection of collaborators is not based on preferential attachment, but more of an assortativity effect creating not merely a rich-gets-richer effect but an elitist network with high entry barriers. In this acclaimed democratic and collaborative environment of Big Science, the elite closes in on itself. We propose this tendency to be even more explicit in other knowledge producing environments with more visible boundaries and higher thresholds for collaboration.

Jelcodes:O39,-

Big Egos in Big Science?

Unlocking Peer and Status Effects in the evolution of collaborative networks

Abstract

In this paper we investigate the micro-mechanisms governing structural evolution and performance of scientific collaboration. Scientific discovery tends not to be lead by so called lone 'stars', or big egos, but instead by collaboration among groups of researchers, from a multitude of institutions and locations, having a diverse knowledge set and capable of tackling more and more complex problems. This prose the question if Big Egos continues to dominate in this rising paradigm of big science. Using a dataset consisting of full bibliometric coverage from a Large Scale Research Facility, we utilize a stochastic actor oriented model (SAOM) to analyze both network endogeneous mechanisms and individual agency driving the collaboration network and further if being a Big Ego in Big Science translates to increasing performance. Our findings suggest that the selection of collaborators is not based on preferential attachment, but more of an assortativity effect creating not merely a rich-gets-richer effect but an elitist network with high entry barriers. In this acclaimed democratic and collaborative environment of Big Science, the elite closes in on itself. We propose this tendency to be even more explicit in other knowledge producing environments with more visible boundaries and higher thresholds for collaboration.

INTRODUCTION

Many great discoveries are results of team effort and the significance of team composition and collaboration has long been celebrated, but the antecedents to collaboration and their relation to collaboration effects remain unexplored. Research continues to demonstrate that creativity of all sorts predominantly stems from collaboration: Collaboration allows individuals to join resources and tackle problems too big for each of them, and through combination of their respective abilities and expertise, individuals can tackle problems too complex for each of them alone. Collaboration in knowledge production has evidently been increasing since the mid 1950es, and most knowledge production, such as papers and patents, are now developed in collaborative teams (Wutchy, Jones & Uzzi, 2007). Collaborative projects have more impact than individual research (Wutchy et.al., 2007), with collaborations spanning organizational boundaries presenting the highest average impact (Jones, Wutchy & Uzzi, 2008). But we still know little of how network dynamics shape the structure of collaboration and knowledge creation (Phelps, Heidl & Wadhwa, 2012; Ahuja, Soda & Zaheer, 2012).

Even though substantial amounts of literature elaborate on the intricate relationship between performance, knowledge production and the network of individuals, results are mixed and inconclusive (see Phelps, 2012 for an overview). One potential reason for this could be the lack of longitudinal network analyses and of models controlling for the inherent endogeneous networks mechanisms affecting collaboration and performance. Actors choose their collaboration partners, and this in turn influences and constrains their future performance and selection (Baum, Cowan & Jonard, 2010). Thus network structure provides a powerful endogeneous force restricting individual agency. The literature on network evolution and structure is dominated by intent on reproducing the topological form of real world networks (e.g. Erdös & Rényi, 1959; Watts & Strogatz, 1998; Barabási, 2002), and has largely ignored tradition in sociology, psychology and economics for regarding the behavior and characteristics of individuals. Whereas literature on knowledge networks neglects the endogeneity of network structure, literature on network structure evolution disregards individual agency. As a result, essential questions regarding production of knowledge and its relation to collaboration and performance remains to be addressed.

In this paper, we address the question of whether collaboration itself affects the performance of the individuals, or if collaboration is simply a byproduct of high performers being more attractive as collaborators. Second, we study the micro mechanisms governing the choices made by individuals to engage in collaboration. We employ an approach allowing us to model both the structural, nodal and behavioral characteristics in the form of a variant of agent based modelling. Thus we are able to represent network and behavior change as the result of dynamics driven by different endogeneous tendencies and individual level agency. This allows us to analyze the microfoundations of collaborative research, the mechanisms driving the observed patterns of increased collaboration effort and pay-off. We unpack mechanisms behind the observed stratification through a longitudinal study of the dynamics of individual level tie formation. We investigate the governing dynamics of multi-institutional scientific collaborations. To analyze the micro-foundations of research collaboration, we turn to a setting purposefully set up to foster collaboration across organizational, institutional and national borders: Large Scale Research Facilities. We consider LSRF a critical case for testing tendencies in research collaboration as they are designed to facilitate boundary spanning collaboration. Researchers are brought together, regardless of geographical, organizational and institutional distance. Joining the facility is subject to application and approval, and low performing researchers are seldom admitted. Thus, researchers at the facility can rest (relatively) assure of their fellows possessing a decent level of skills.

Furthermore, LSRF are centered on instruments central to experiments, and we should thus expect a relatively low knowledge spread. Qualitative studies of LSRF show one purpose of joining these facilities is to be collaborative and to meet new collaborators from different institutions and countries. In this literature, team formation is described as rather democratic – i.e. not stratifying but cross-organizational and inter-disciplinary. Consequently, we chose a LSRF, more specifically the Spallation Neutron Source (SNS) and the High Flux Isotope Reactor (HFIR) located at Oak Ridge National Laboratories, Tennessee as a critical case on which to test and further separate, the effects of network endogeneous mechanisms and individual agency. If big egos exist here, we should expect them to be predominant throughout many (scientific) fields and settings. We find that especially the mechanisms of embeddedness, status and homophily (similarity in activity and abilities) are main drivers of establishing team work. These individual level mechanisms are the microfoundations for the creation of the inter-connected elite of high performing researchers found in Jones et al. (2008). Second, we find no observed status effect on performance, but instead a high significance of peer influence. This adds to the body of literature not only on scientific collaboration, but also extends theory on knowledge networks through the linking of causal mechanisms of actor's network position, and local network structure to the network evolution and performance of individual scientists.

THEORETICAL BACKGROUND

Collaborative research has become the model per se in many fields of science due to specific benefits: *First,* it facilitates division of work and the pooling of intellectual expertise. *Second,* collaborating permits the accomplishment of projects that could not be realized by a lone scientist (Katz & Martin, 1997). This can especially be seen in interdisciplinary studies or research involving specific instrumentation (Chompalov, Genuth & Shrum, 2002). Collaboration increases the number of studies that can be undertaken and sharing the workload increases the number of projects each individual can embark upon. Therefore, collaboration increases the probability that an author will see some parts of his/her work accepted for publication in a journal (Barnett, Ault & Kaserman, 1988). Co-authored papers also present a higher quality than those which are single-authored (Laband and Tollison, 2000), which leads to a higher average impact (Wutchy et.al., 2007; Katz and Martin, 1997). The motivation for scientists to join forces and collaborate can thus be seen as a product of necessity – in the form of division of labor, and access to relevant knowledge and resources – and as the anticipation of added output, impact and quality.

Next comes the question with whom to collaborate, in this paper we analyze how the mechanisms of embeddedness, homophily and status guide choice of collaborators. From a relational perspective one of the selection mechanisms governing the choice of collaborators has long been ascribed to be contingent on the network of ties in which people are embedded (Simmel [1922], 1955). Embeddedness increases opportunities for social sanctions and thereby mitigates opportunism and conflict (Granovetter, 1985; Uzzi, 1997). Nothing reassures as the recommendation of a potential collaborator by a trusted third party (Burt & Knez, 1995;), similarity on observable dimensions also inspires trust and favorable assessments (McPherson et.al. 2001), and collaborators are often chosen based on homophily on observable dimensions (Hinds, Carley & Krackhardt, 2000; Aarts and Dijksterhuis, 2000; McPherson, Smith-Lovin & Cook, 2001). Embeddedness and homophily both decrease uncertainty and establish trustworthiness of potential collaborators (Granovetter, 1985), low uncertainty and high trustworthiness in turn increase probability of successful collaboration. The tendency to rely on embeddedness and homophily in choice of collaboration partners based on the expectation that such partners will prove superior, can thus serve as a self-fulfilling prophecy (Sorenson & Waguespack, 2006). Another common denominator found to impact the choice of collaborators is that of status. High ability level collaborators are generally desirable as peer effects improve the joint output of the collaboration. Studies of peer effects in parallel execution of routine tasks (Falk and Ichino 2006) as well as in more complex team production settings (Bandiera, Barankay & Rasul, 2010a;2010b) illustrates that peers influence norm establishment and compliance levels and through these mechanisms affect performance levels. But as information on ability levels is seldom readily available, agents must rely on signals such as collaboration activity level and past performance.

Previous studies has thus presented evidence of the individual motivations guiding collaboration, and in this paper we address how these motivations are embedded in an evolutionary process of scientists collaborating in scientific networks. The importance of scientific networks can hardly be overstated. Extensive literature has dealt with the role of knowledge networks in science, research and innovation. Knowledge diversity facilitate innovation through recombination (Galunic & Rodan, 1998), and positions in the knowledge network as well as the number of co-authors affect scientific production and impact (Wuchty et.al., 2007; Abbasi et.al., 2011). Distance in terms of knowledge fields, organizational or institutional affiliation have been shown to both impede collaboration and improve collaboration outcomes. For low innovation level projects, the challenge of integrating different perspective is disruptive, while highly innovative projects benefit from the diversity of knowledge to recombine (Bercovitz & Feldman, 2011). A common factor characterizing these studies is an emphasis on how knowledge networks facilitates the creation, transfer or adoption of knowledge (Phelps et.al., 2012).

Less attention has been paid to understanding the endogenous nature of network emergence and evolution, and how it affects performance (Ahuja, Soda, & Zaheer, 2012). Thus, recently, a literature stream has started to emerge with focus on using the dynamics of the network to explain collaboration and performance. This research especially address the macro-dynamics of networks e.g. how organizational fields evolve or the extent to which knowledge flows are geographically mediated and is driven by concurrent forces operating at the micro level (Powell, White, Koput & Owen-Smith, 2005; Azouley, Zivin & Sampat, 2012). The focus on micro mechanisms is especially important because agents usually are not able to cast their gaze across the entire network at one time. From their localized perspective, they form ties and make decision based upon the intersection with those that are socially proximate, socially visible or where they can tap into the specific human capital that comes from collaborating with high perfoming alters (Robins et.al., 2005; Azoley, Zivin & Wang, 2010). Thus, individual level perception affect network structure and network structure affect the individual level perception. Analyzing such microfoundations, while accounting for both exogeneous characteristics and endogeneous network evolution, is consequently pivotal for understanding collaboration choices and collaborative outcomes.

In the following sections we will present hypotheses related to agency driven as well as network endogeneous mechanisms.

Revisiting the Matthew Effect

"The world is peculiar in this matter of how it gives credit. It tends to give the credit to already famous people." (statement by Nobel laureate in physics, Zuckerman, 1965, in Merton, 1988). This mechanism of social agents in which a favorable relative position becomes a resource that produces further relative gains, was originally explained and termed by

Robert Merton in 1968 as a means to explain variation in the advancement of scientists. 'Nicknamed' the Matthew Effect, this mechanism has been shown to have general applicability for explaining the emergence and increase of inequality across many temporal processes (DiPrete & Eirich, 2006). In general, the literature distinguishes between two sub effects: Preferential attachment – the social status effect of highly visible agents to become increasingly centralized in a network of agents - and cumulative advantage – the positive effect of high social status on performance.

Preferential attachment refers to a network endogeneous tendency for each node to have higher probability of – i.e. to prefer – to form linkages – i.e. attachment – with prominent alters. On the nodal level, preferential attachment has been identified as a governing mechanism for collaborative choice, both for individuals and firms. It has been argued to guide many aspects of human behavior, from location choices of human capital (Lorenzen & Andersen, 2009; Krugman, 1996; Christäller, 1933), to internet browsing (Barabási & Albert, 1999), and choice of collaboration partners (Newman, 2002). On the structural level, the preferential attachment mechanism contributes to highly unequal network structures with increasing inequality among nodes, and it has been found to hold across settings and analysis levels. In their seminal paper, Barabási & Albert (1999) use the notion of preferential attachment in their mathematical modeling of graph evolution, finding a large correlation with real world networks, and thus explaining the scale-free networks usually found in both collaboration and information, e.g. citation networks.

The tendency for preferential attachment in ties formation has a counterpart in productivity. The distribution of productivity among scientists is also highly skewed with initially highly productive scientists maintaining or increasing their productivity while scientist who produces very little produce even less later on (Allison & Stewart, 1984). We content to see the preferential attachment mechanism as a result of agents perceiving many connections to be a signal of high social status. Consequently, the effect of preferential attachment becomes an increase in influx of opportunities to well-connected agents. This mechanism addresses the antecedents of collaboration: an ex ante preference to connect with those already rich in collaboration partners, but it does not address ex post performance effects. Regardless of whether the increase in influx of opportunities to well-connected agents is justified in real improved abilities or rests on irrational perceptions, the consequences remain to be that well-connected agents experience improved opportunities to excel - an ex post effect translating status (many collaborators) to improved performance. The same reasoning can be found in Merton's (1968) original work, where status not only influences the perception of quality, but scientists having a high status are more likely to attract both tangible and intangible resources, which in turn can result in scientific outputs of higher quality. This transformation of social status to increased performance we term cumulative advantage, and can be termed as effect of increasing an agent's centrality in the network on performance.

Based on the network endogenous mechanism of preferential attachment and cumulative advantage, we then propose:

Hypothesis 1a. Agents having many collaborative partners experience increased probability of forming new ties.

Hypothesis 1b. Agents having many collaborative partners experience a positive impact on their performance.

Collaboration with Stars

Above and beyond network endogeneous mechanisms, agents will tend to seek out collaboration partners with prominent alters, meaning that all agents irrelevant of performance level will prefer to connect with alters performing at higher levels than themselves. Previous research shows that individuals with prior knowledge or indication of potential collaborations partners' competence levels prefer to collaborate with individuals with competence at the same or higher level as their own (Schwab, 2009; Skilton, 2008). This is cause by two interrelated dynamics. First, if agents are connected to prominent alters, it sends a signal of high quality to third parties who are impressed by the "reflected glory" expressed by the tie (Cialdini et al, 1976; Kilduff and Krackhardt, 1994). This can lead agents to prefer ties to prominent alters based on an expectation of spill-over effects from the connection. Second, to reach and maintain high levels of performance, highly skilled individuals need to collaborate with other highly skilled individuals (Groysberg, Lee & Nanda, 2008; Groysberg, Lee & Abrahams, 2010). Less capable collaboration partners will reduce the overall performance level of the collaboration outcome, while more capable collaboration partners will increase it. Consequently, it is rational for ambitious agents to seek out alters at higher performance levels.

Hypothesis 2. Agents will seek collaborate with alters performing at higher levels than themselves.

Birds of a Feather

Another mechanism exogeneous to network structure is an inherent tendency for agents to prefer collaboration with alters similar to them self on relevant and observable dimensions. Also known as homophily, this mechanism origins from sociology and addresses ex ante selection. Selection based on homophily often guides collaboration in settings with many unknown agents and lack of complete information on peer capabilities (Hinds et al., 2000; McPherson et al., 2001). Thus homophily strongly affects formation of collaborations as agents prefer similar alters and a priori expect self-similar alters to be more likely to both accept proposals and prove beneficial collaboration partners. In our chosen setting, similarity in organizational affiliation, research field, research activity and quality could be guiding an agent's propensity to prefer collaboration with alters, but we focus on the aspects most directly relating to the social status mechanisms identified earlier: social visibility in terms of collaborative activity and performance.

Agents are capable of capturing signals related to alters' status and regard those in similar structural positions as peers, and will strive to imitate their professional behavior (Burt, 1987). Agents who are more structurally similar to one another are more likely to have increased interpersonal communication (for a prominent example of this research see Burt, 1982), and we can hence expect agents to seek out collaboration partners based on similarity in network position – i.e. in activity level. Highly active agents will seek out highly active collaboration partners. This mechanism can be extended to go beyond mere activity and also be relevant for performance. Empirical findings on social networks show this to result in highly skewing effects (Newman, 2002).

When agents sort collaboration partners' based on capability levels – using social status and quality as signals - prominent agents tends to choose collaborating with each other (Newman, 2002). To tease out the extend of this effect, i.e. whether collaboration is caused by homophily related to similarity in social status or a result of similarity in performance, we pose:

Hypothesis 3a. Highly active agents tend to form new collaborations with other highly active agents.

Hypothesis 3b. Agents tend to form new collaborations with other agents performing at similar level.

Once collaboration is established, individuals' behavior tends to affect and be affected by social interactions. Deviant behavior, opinion formation, generation of research ideas and the emergence of status hierarchies are diverse examples in which peer effect are likely to play a role (Lomi, Snijders, Steglich & Torló 2001; Rawlings & McFarland, 2011). Identification of peer effects in parallel execution of routine tasks (Falk and Ichino 2006) as well as in more complex team production settings (Bandiera, Barankay et al. 2010; Bandiera & Barankay. 2010) illustrates that peers do influence norm establishment and compliance levels and through these mechanisms affect performance levels. Results show that peer established norms induce low-productivity workers to increase effort, while high-productivity workers are willing to reduce effort and forego earnings to comply with established norms (Falk and Ichino 2006; Bandiera, Barankay et al. 2010; Bandiera, Barankay et al. 2010). Scientists establish collaboration to divide labor, access knowledge and improve the overall quality of their work. Thus, if selecting collaborators are a deliberate matching choice based on signals of performance and status, the consequence of peer effects is that agents will tend to collaborate at similar levels of performance. However, if the matching procedure are only partially deliberate and based upon other social factors than quality, this will result in biased estimates of peer effects (Azouley, Ding & Stuart, 2009). Thus, we hypothesize:

Hypothesis 3c. Agents tend to assimilate to their collaborative partners level of performance, controlling for network endogeneous mechanisms.

EMPIRICAL SETTING

Big Science and Large Scale Research Facilities

Big Science requires big budgets, big planning and big collaborative effort. The trade-off for these big time investments are the potential for breakthrough discoveries, both in the scientific world and as spillovers in the form of inventions with radical potential. When research involves site specific, large and complex instrumentation, as is the case with Big Science, collaboration is especially common (Katz & Martin, 1997). Due to the complexity of using different instrumentation and diverse knowledge skillsets necessary to be able to analyze the output, co-authoring and collaboration at these sites can be thought attributed to necessity as well as to intellectual overlapping or spontaneous meetings. Accordingly when the knowledge base of a research project is characterized by a high level of complexity and dispersed pool of expertise, the locus of innovation will be more likely centered in collaborative networks (Powell et. al., 1996).

We choose the setting of a Large Scale Research Facility, because it provides us with a geographical localized multi-institutional context, with distinct roles assigned to scientists, according to e.g. the instruments they are operating or whether they are residents or visiting scientists. At the same time a facility like this serves as an extreme case of the paradigm change and professionalization connected with the rise of big science. Some articles have focused on these sites, but has thus far either delved with the learning perspectives of the individual (Boisot et.al., 2011; Autio et.al.. 2003), drivers of internationalization (Lauto & Valentin, 2013) or various case studies focusing on the different 'spillover' effects (Langford & Langford, 2000; Merz & Biniok, 2010). Research in multi-institutional collaboration in the natural sciences has been primarily dominated by historians, sociologists and anthropologists, focusing in particular on high-energy particle physics (Chompalov et.al., 2002). This has provided an excellent, but disproportionate view on collaboratives in big science as *"post-traditional communitarian formations with object-centered management, collective consciousness, and decentralized authority"* (Ibid., p. 751). This has been described as an example of the new model for collaboration in science (Knorr Cetina, 1999), but also been challenged. Chompalov et. al. (2002) show that this mode of organizing in multi-institutional research projects is the exception of the rule, and is predominantly found in the HEPP community. These facilities have a high influx of highly skilled people from all over the world, from a multitude of institutions, successfully facilitating the establishment of ties between distant organizations.

The research setting

As our empirical setting we choose the Spallation Neutron Source (SNS) and the High Flux Isotope Reactor (HFIR) located at Oak Ridge National Laboratories (ORNL), Tennessee. Established in 1943, the overall facility is a multidisciplinary center financed solely by six members of the U.S. Department of Energy. The facility conducts both basic and applied science in specifically the areas of neutron science, biological system, energy and high energy physics, advanced materials, supercomputing and national security. Approximately 4,600 scientists are employed at Oakridge and the facility had a budget of USD 1.65 billion in 2011. Since 2006, the research program in neutron science is managed by the Neutron Sciences Directorate. Our chosen setting ORNL/NSD employs 600 scientists, technicians, and administrative staff and operates two of the world's most advanced neutron scattering facilities: a Spallation Neutron Source (SNS), which became operative in 2006, and a High Flux Isotope Reactor (HFIR), completed in 1965 and renovated in 2007. In our study we focus on the knowledge production surrounding these two facilities.

DATA AND METHOD

Collaborations evolve within social spaces comprised of a complex interlocking of socio-demographic, organizational and intellectual factors each of which pushes and pulls researchers toward interacting with specific individuals. One simple, but powerful, indicator of collaboration is the co-authoring of an article. This formal tie may not capture all interaction, but it serves as a lower bound for social interaction (Singh, 2005). We use a stochastic actor oriented model (SAOM) to analyze network evolution of scientific collaborations, incorporating both structural and behavioral effects in a longitudinal perspective. The approach utilized in this paper thereby contributes to an active research domain, which seeks to disentangle social selection from influence (Snijders et.al., 2007; Steglich et.al. 2010; Giuliani, 2014; Ter Wal & Boschma, 2009), and draws upon recent statistical advances in the network literature to model such processes with greater confidence (Snijders 2001; Steglich et.al. 2010).

Data

The empirical study investigates the evolution of scientific collaboration in the context of SNS & HFIR. Since 2006, all peer-reviewed publications based on research utilizing SNS & HFIR data and resources, or conducted by staff affiliated with SNS & HFIR are publicly listed on the directorate's website. We refer to these publications as SMS & HFIR-based research. We retrieved full bibliometric records from ISI-Web of Science of the publications produced in the period from 2006 to 2009. A total of 2180 distinct scientists stemming from 687 publications where collected (data collection ongoing).

Modeling Dynamic Networks

To model the temporal dynamics of networks at the LSRF, we apply a stochastic actor-based approach, which can estimate the evolution of social networks, in terms of tie establishment between agents, as driven by exogenous as well as endogenous forces. In detail that means the probabilities of tie changes is modeled as a function of individual actor characteristics as well as their network position. It enables us to capture endogenous effects of high importance when explaining the evolution of social networks. Specifically we utilize the statistical framework for analyzing network panel data called SIENA (Snijders et.al., 2010). This model combines random utility models, Markov processes and simulation to estimate network parameters between observations and concurrently estimate parameters for the underlying network dynamics. SIENA can be categorized as a combination of mathematical models and traditional econometric models. Even though this is a powerful analytic tool, four fundamental underlying assumptions has to be met (see also Snijders et al., 2010) First, ties must reflect an association agreed upon by both agents involved; second, only the current state of the network affects its future evolution; third, ties must represent an enduring relation; and fourth, changes must be seen as a stepwise process.

Key Variables

Dependent variables

The modeling in the SIENA-methodology employed in this paper is given by the network and behavioral dynamics which are subjects to change almost simultaneously – this means the behavioral dynamics change dynamically in mutual dependence with the changing network.

The dependent variable of the network change consists of the parameter estimates influence on an agent's decision to establish new ties – basically modelling the 'attractiveness' of a given network configuration for a given agent. E.g. we model the probability of forming a tie.

The dependent variable of behavioral – i.e. performance – change consist of the log-odds-ratio of an increase in behavior compared to staying constant. As an example, when comparing two scientists who are equal in all aspects, except that the collaborators of the first on average perform better by 1 unit, on the performance scale, the odds of increasing ego's performance (compared to no change at all) are given as the exponentiated parameter estimate. Due to the complexity and time-consuming nature of the model, when forward citations are used as dependent variable, we transform it into a categorical variable separated by quartiles. Following we end up with a dependent variable for performance in the simulation model having four distinct levels.

Independent variables

The individual parameters available in SIENA can broadly be divided into three categories: (i.) Network base effects, referring to the agents general tendencies to form ties in a particular way, independent of alter and ego's network position and other characteristics. (ii.) Degree related effects capturing the endogenous influence of several effects associated with alter and ego's degree of ties. (iii.) Covariates which are exogenous characteristics of the actor. These variables are all included in both the objective function and the performance part of the models. In the following we discuss in detail the main effects included in the model.

"Preferential Attachment" represents the tendency of agents to form ties to alters already receiving a high degree, hence popular ones. A positive estimate implies a self-reinforcing mechanism that over time leads to increasing dispersion of the degree distribution of the networks (Newman, 2000).

"Cumulative Advantage" represents the effect of actors having high social status in the form of collaborators to also affect their performance. The measurement of degree as cumulative advantage follows (Dahlander & McFarland, 2013).

"Similar activity level" refers to the preference of agents to form ties with alters based on their own as well as alters degree[1]. This combination represent a measurement for homophily as well as social stratification in the network pattern, and in the case of the LSRF that scientists having a high social status team up with scientists with similar social status.

"Similar performance" is a dyadic transformation of citations defined in such a way that it is scaled between 0 and 1, with 0 meaning that one author has the minimum value of citations and the other has the maximum (maximum dissimilarity), and 1 meaning that two authors has the same citations (maximum similarity).

"Performance alter" refers to the general tendency of agents to choose to collaborate with alters performing at a higher level than themselves.

Controls

We include a number of controls to account for the possibility of spuriousness, and alternative sources of influence and selection.

[1] In the network literature this is often called the "assortativity"-effect.

“Triadic closure” is an effect most commonly found to drive network evolution. It describes the tendency of a triad to be closed, i.e. if *i* are friends with *j* and k, what is the probability that j and k also will be friends. Basically it tells something about the degree of clustering in the network. Thereby we show whether partners partners in year t are more likely to collaborate in time t+1.

We control for average influence of citations on degree by incorporating the interaction term “Citations x Ego”.

“Same organization” and “same country” are binary variables indicating whether or not the scientists stem from the same institution or from the same country.

“Organization type controls” are a list of binary covarying variables indicating whether or not scientists are of the same institutional type, based on the following categories: 1) Resident at the SNS, 2) University, 3) Research Lab and 4) Business. We include these in our models to control for the possibility of collaboration being instigated by the specific division of work at LSRF’s, e.g. an instrumental scientists employed at the facility.

To control for the fact that both collaboration and performance can be driven by the specific cognitive profile of each focal scientist, we rely on the classification developed by CHI research (Narin, Pinski & Gee, 1976; Lauto & Valentin, 2013). The classification scheme utilized thus assigns ISI-recorded journals to a research level on basis of content and scientific field. For non-biomedical fields, the values are 1- “applied technology”; 2 – “engineering”; 3 – “applied research-targeted basic research”; 4 – “basic scientific research”.

Table A1 depicts the complete list of variables included in the model:

Insert Table A1 about here

For the SAOM we exclude from the network scientists not part of the final largest component. This is due to the fact that these 'one timers' would potentially result in an overestimation of network measurements such as e.g. triadic closure similarity and activity.

DESCRIPTIVE STATISTICS

Insert Table A2 about here

Table A2 shows the distribution of scientists according to country of origin. The majority of scientists are stemming from the US, seconded by Germany, the Japan and finally France. This relationship is not surprising, i.e. Germany and France has two of the largest neutron scattering facilities in Europe (Helmholtz-Zentrum Berlin & Institut Laue Langevin), and the same is true for Japan, where a similar facility has been constructed at JPARC. This highly skewed distribution of scientists in favor of the US (~60%), seems to give indications of the SNS giving priority to employ local scientists, which could potentially bias our variable of same country. The outcome effect of the variable on the network selection model would overemphasize the effect on connectivity due to the fact that the mere presence of US based scientists increases the likelihood of a random encounter, and further the number of scientists stemming from other countries could be due to a strategic decision made by the top of the board at the facility.

Insert Table A3 about here

Table A3 shows the different metrics of important variables employed in this study. The average collaboration employs 6.7 scientists, with a standard deviation of 5.80. Even though the largest team is 97, the fact that the standard deviation is smaller than the mean indicates that the problem of very large team structures, as seen at work at other LSRF's are not an issue of concern here. The mean number of publications pr. scientist is 2.47 and the standard deviation is 3.40. The maximum number of publications by a scientist is 35. This results in a skewed distribution of scientist productivity, meaning the existence of potentially some "star scientists" with a high productivity rate, and a long tail of low producing scientists. Approximately 65% of the scientists are only producing a single publication associated with the facility, resulting in a potentially biased estimation of publication related variables e.g. performance and CHI, as these variables thus are only based on one observation of the scientist. Therefore we employed a robustness check where we separated the scientists having only one publication, and the rest, and ran two distinct models for each, with similar results on the network selection part.[3] For the performance variable we observe a similar skewed distribution, with a few stars getting a lot of citations and many getting 2 or fewer

----

[3] We are currently working on collecting data on scientists facility external publication history before and during their time at the SNS, which, when employed, would raise the validity of our models.

Insert Table A4 about here

Insert Table A5 about here

Table A5 and A6 show the overall evolutionary network statistics. Due to the fact that only upward movements are allowed (i.e. no dissolution of ties), interpretation should be made with caution, but we can see that the average degree of each scientist starts of at approximately 9 and ends at almost 17, indicating a highly collaborative environment. The change in network ties is especially present in the period 3→4, where a little over 14,000 new ties where made. We can see that the Jaccard coefficient - indicating the degree of change between network observations - are high. Thus estimation by SIENA seems appropriate.

RESULTS

Table A6a and b shows the results of the stochastic actor-based analysis.

Insert Table A7a about here

All parameter estimations are based on 1,000 simulation runs. After a series of iterations the approximation algorithm converged somewhat excellent for all variables (all around values 0.1). The convergence indicates whether the deviation of the simulated structures compared to the observed structures is acceptable (t-values < 0.1), and can be used as an approximate goodness of fit of the different parameters. Similar activity and preferential attachment was not able to run joined in the same model due to issues of multicolinearity. The parameter estimates can be interpreted as non-standardized coefficients obtained from logistic regression analysis (Steglich et al. 2010). Therefore, the parameter estimates that are reported can be read as log-odds ratio, i.e. how the log-odds of tie formation change with one unit change in the corresponding independent variable, and how the log-odds of increasing performance by one unit with a one unit change in the corresponding independent variable. Odds ratio can be computed as the exponentiated form of the coefficients of each predictor.

As opposed to our hypothesis (H1a), the measurement for preferential attachment, is significant but surprisingly with a negative coefficient (exp(-0.17)=0.84), meaning when increasing the number of collaborators above the mean, scientists are less likely to establish collaboration with you. Instead they tend to regress around the mean. For H2 we find no statistically significant indications that scientists are more likely to establish ties with scientists with a higher count of citations. At the same time though, the significant and positive effect of both similar activity (H3a) (exp(0.4852)=1.62) and similar performance (H3b) (exp(1.3202)=3.74 & exp(1.5451)=4.69) gives indication of the existence of a highly elitist network effect where the scientists collaborating the most predominantly chooses to collaborate with others of similar social status and visibility. Our controls on especially whether scientists have the same organization (exp(0.62)=1.86) or are stemming from the same country (exp(0.70)=2.01) yields significant and positive estimates, showing that even though we observe a high degree of multi-institutional collaboration, geographic proximity are still a main mechanism leading to collaborative ties at this LSRF. As for the controls concerning the cognitive profile of scientists, we find that the more basic oriented a scientist is, the higher the degree of collaboration they exhibit (exp(1.4868)=4.42), much in line with the findings of (Lauto & Valentin, 2013).

For the performance dynamics we find that the rate functions – describing the average number of opportunities for change for each scientist between observations – decline over time. Thus the opportunities peaks in the first period and then sharply declines in the second and third. This suggests that the performance of the scientists tend to stabilize over time. Results from the performance average similarity (H3c) suggests a systematic existence of peer effects, meaning that the performance of individual scientists tend to assimilate to the performance of his or her peers. The effect of cumulative advantage (1b) is negative but remains insignificant in both models.

The significance of the quadratic shape indicates a performance effect's effect on itself, showing it becoming increasingly difficult to perform better, i.e. the effort of increasing performance from 3-4 is larger than 2-3.

DISCUSSION

One of the peculiar findings in this paper is the observed negative effect of preferential attachment. In effect this may be a reflection of the apparent status hierarchy enacted in the network. Our models show high performance to lead to more collaboration partners, and with many collaboration partners come the opportunity to repeat interactions, estimate to reap the most benefits. Thus, high performing agents have better opportunities for selecting collaboration partners. Following, high performing scientists have increased opportunities for making use of the advantage of lasting ties. The costs of establishing ties and the persistence of said ties vary (Dahlander & McFarland, 2013). Establishment of ties hinges on complementarity and opportunity recognition across unfamiliar partners, while tie persistence depends on the experience and outcome of the collaboration (Dahlander & McFarland, 2013). The establishment costs outweigh persistence costs and risks are lower in the reusing of ties,, but new ties provide the benefit of fresh perspectives and complementary knowledge which is assessed to (and sometimes do) balance the costs. These differences would benefit high performing agents more. If we assume that benefits associated with reusing an already familiar collaboration partner decrease at increasing rate, highly connected scientists have more collaborators with whom they can engage in repeat collaboration before reaching a suboptimal outcome. They can thus enjoy the low costs of tie persistence without endeavoring into a level of relational embeddedness mitigating the benefits of complementarity too much. Whether these mechanisms rest on an initial superior ability to select collaboration partners is unfortunately beyond the scope of our analyses.

With repeat interaction, the intensity of ties increase and we should expect tie strength to vary across one-off and repeated collaborations. However, context, setting and product remains the same, and thus, benefits of collaboration could vary with tie strength. Normally, we would expect tie maintenance cost to increase with tie strength (and the number of ties maintained) (Granovetter, 1973), but in our case, ties and their purpose remains the same across tie strength and hence cost should remain the same or even decrease. We would also expect uncertainty to decrease and trust to increase with increasing tie strength. Further, the propensity for triadic closure could reasonable be argued to increase with tie strength. Consequently, opportunity and recommendation would lead to network cohesion. In our setting of scientists collaborating within a large scale research facility, this would further the hierarchical mode of collaboration leading to a network with a distinct elite of highly active and high performing scientists collaborating amongst themselves, disregarding lower status and lower performing scientists. Unfortunately, one weakness of the SIENA method is that we cannot distinguish between one off and repeat interaction. Hence, we are left to infer from our findings on this topic.

Delving further into the mechanisms why our findings do not reflect the positive preferential attachment and cumulative advantage effects found in previous work we propose the explanation to be our method of taking network endogeneous selection mechanisms into account. Consequently, our other key variables capture much of the variation typically labelled as preferential attachment and cumulative advantage. The similar activity, or assortativity mechanism, leads to collaboration among scientists with similar collaboration activity and performance and gives rise to a "rich old boys club" of elite scientists We presume these positive effects to cause the negative effect of preferential attachment: when assortativity guides establishment of new collaboration ties, it is difficult for the lower performing scientists to engage stars, or big egos, (social and performance wise) in collaboration.. These mechanisms guide selection into a hierarchical system where collaboration among the most prominent agents with respect to network position and previous performance becomes relatively closed to agents of lower performance. Thus, the effect of cumulative advantage for performance is reduced to insignificance by agents selecting and rejecting collaborators in a highly stratified system of collaboration.

Taken together our results suggest that connecting is not only a function of performance, but also a social process focusing on status and governing the evolution of the entire network. This results in a stratified network structure with highly active scientists steering network evolution. These findings are much in line with the findings of Jones et.al. (2008), and thus contributes with understanding of the microfoundations for evolution of the networks observed throughout this literature stream. But more research is needed to tease out the effects of actor knowledge, and knowledge overlaps on network evolution, i.e. whether the observed stratification comes at the cost of knowledge heterogeneity, and finally how this influences individual performance.

The creation of ties at the Spallation Neutron Source is to a large degree governed by social hierarchy. Scientists with large collaborative capacity choose to collaborate with other scientists with similar capacity. This creates a highly elitist closed and separated network, which is surprising given the literature on Big Science and the 'democratizing' scientific research paradigm anticipated here. The findings serve as micro level explanation for the results found by Wuchty, Jones and Uzzi (2007; 2008), who found development of scientific collaboration to be increasingly elitists governed. The significantly positive impact of assortativity tells us, that in the observed collaboration network, scientists tend to collaborate with others at similar levels both with respect to productivity, connectivity and performance, and thus that joining the elite collaborators is hard and becomes increasingly hard for the average scientist. This should not necessarily be seen as unfortunate. An assortative network tends to percolate more easily, creating a giant component faster than a disassortative. Thus high-degree nodes will tend to stick together in the form of a core group, making dissemination of knowledge happen faster but at the cost of the resulting size of the giant component (Newman, 2002; 2004). This type of assortative network is typical to social settings and has been shown to lead to fast spread of ideas, but also to rapidly increasing similarity and very robust network structure where removing a few collaborators has little impact on the structure (Newman, 2002). This would be the type of network, where we could expect social and human capital to correlate and co-evolve (Coleman, 1988) giving rise to paradigmatic worldviews (Dosie, 1982). While working within a paradigm can be productive, the robust structure of this type of network will hinder paradigmatic changes when needed. However, one of the arguments for centralizing scientific investigation and collaboration at places such as the SNS are funded upon an innate ability to facilitate inter-disciplinary collaboration, these findings suggests that the localization and structural features of LSRF are not enough to facilitate this meld. Instead the formation of the core group of highly active researchers, and the increased probability of forming ties based on similarity in social status could even mean less inter-disciplinarity.

Future research

The hierarchical structure of networks should be studied with great attention to the role and functions of the agents inhabiting these networks. In our paper, function and affiliation were reduced to controls, but in the context of the specific facility, some scientists are experts on specific instruments, this function may lead to specific co-authoring patterns, potentially making them the hubs of the Big Science world. Thus further research should be conducted on the gatekeeping roles of specific scientists in the world of big science. Further, future research could contribute to our understanding of network evolution and resistance by analyzing one-off versus repeat collaboration in order to answer the question raised in this paper of whether high performers utilize their large network to repeat collaboration. Thus, further research should delve into the mechanisms governing the selection of partners, more than the mere structural observation of degree.

CONCLUSION

In this paper we asked the questions what the governing dynamics driving scientific collaboration in modern day science are, and what separates high performing scientists from lesser so. Drawing from mostly empirical research on scientific collaboration, we show that literature pointed towards a paradigmatic change in the conducting of science, placing more and more emphasis on the connective and collaborative capacity of scientists to team up and cross organizational boundaries. We further highlight the need for a longitudinal perspective in order to address the influence of network evolution in scientific collaboration. We propose the hypotheses that science is increasingly driven by not only preferential attachment, but also by notions of proximity and homophily, with a special emphasis on the assortativity effect as a driver of the "rich old boys club"-effect. We investigate this in the form of bibliometrical analysis of the empirical setting of large scale research facilities, more precisely located at Oak Ridge National Laboratories in Tennessee. We conduct the analysis based on a selection of the full bibliometric recording, from 2006-09, of publications affiliated with the Spallation Neutron Source and the High Flux Isotope Reactor. To test our hypotheses we employ a stochastic actor based network analysis to separate the selection and network dynamics from the behavior of external collaboration. By doing so we are able to analyze the cumulative and self-reinforcing effects of network dynamics. We find that we indeed see a network dominated by assortativity and homophily in choice of collaboration partners and by collaboration partner selection and peer effects as main drivers of performance. A tendency of not going beyond your place and collaborating with scientist above your own level, combined with peer effects indicate both a highly unequal distribution of collaboration but also one dominated by hierarchy.

REFERENCES

Aarts, H., & Dijksterhuis, A. (2000). Habits as knowledge structures: automaticity in goal-directed behavior. *Journal of personality and social psychology*, *78*(1), 53.

Abbasi, A., Hossain, L., Uddin, S., & Rasmussen, K. (2011). Evolutionary dynamics of scientific collaboration network. *Scientometrics, 89* , 687-710.

Adams, J. (2004). Industrial R&D Laboratories: Windows on Black Boxes? *The Journal of Technological Transfer, 30* , 129-137.

Ahuja, G., Soda, G., & Zaheer, A. (2012). The Genesis and Dynamics of Organizational Networks. *Organization Science, 23* , 434-448.

Allision, P. D., & Stewart, J. A. (1974). PRODUCTIVITY DIFFERENCES AMONG SCIENTISTS: Evidence from Cumulative Advantage. *American Sociological Review, 39* , 596-606.

Autio, E., Hameri, A., & Vuola, O. (2004). A framework of industrial knowledge spillovers in big-science centers. *Research Policy, 33* , 107-126.

Azoulay, P., Zivin, J. S., & Sampat, B. N. (2012). The Diffusion of Knowledge Across Time and Space. In J. Lerner, & S. Stern, *The Rate and Direction of Inventive Activity Revisited* (pp. 107-155). Chicago: The University of Chicago Press.

Azoulay, P., Graff Zivin, J., & Wang, J. (2010). Superstar extinction. *Quarterly Journal of Economics*, *25*, 549-589.

Azoulay, P., Ding, W., & Stuart, T. (2009). THE IMPACT OF ACADEMIC PATENTING ON THE RATE, QUALITY AND DIRECTION OF (PUBLIC) RESEARCH OUTPUT. *The Journal of Industrial Economics*, *57*(4), 637-676.

Barabási, A. E. (2002). Evolution of the social network of scientific collaboration. *Physica A 311* , 590-614.

Bandiera, O., I. Barankay, et al. (2010). "Social Incentives in the Workplace." Review of Economic Studies 77(2): 417-458.

Bandiera, O., I. Barankay, et al. (2010). Team Incentives: Evidence from a Field Experiment. London, London School of Economics.

Barnett, A. H., Ault, R. W., & Kaserman, D. L. (1988). The Rising Incidence of Co-authorships in Economics: Further Evidence. *The Review of Economics and Statistics, 70* , 539-543.

Baum, J. A., Cowan, R., & Jonard, N. (2010). Network-independent partner selection and the evolution of innovation networks. *Management Science*,*56*(11), 2094-2110.

Beaver, D. d., & Rosen, R. (1978). Studies in Scientific Collaboration, p. 1. *Scientometrics, 1* , 65-84.

Bercovitz, J., & Feldman, M. (2011). The mechanisms of collaboration in inventive teams: Composition, social networks and geography. *Research Policy, 40* , 81-93.

Boisot, M., Nordberg, M., Yami, S., & Nicquevert, B. (2011). *Collisions and Collaboration: The Organization of Learning in the ATLAS Experiment at the LHC.* Oxford: Oxford University Press.

Coleman, J. (1988). Social capital in the creation of human capital. American Journal of Sociology,
94(Suppl.), S95–S120.

Borgatti, S. P., & Halgin, D. S. (2011). On Network Theory. *Organization Science, Articles in Advance* , 1-14.

Boschma, R. (2005). Proximity and Innovation: a critical assessment. *Regional Studies 39(1)* , 61-74.

Bozeman, B. (2000). Technology transfer and public policy: a review of research and theory. *Research Policy, 29* , 627-655.

Bozeman, B., & Crow, M. (1990). The environments of U.S. Laboratories: Political and Market Influences. *Policy Sciences, 23* , 25-56.

Burt, R. (1982). *Towards a structural theory of action: network models of social structure, perception and action.* Academic Press.

Burt, R. & Knez, M. (1995). Kinds of third-party effects on trust. *Rationality and society*, *7*(3), 255-292.

Cairncross, F. (1997). *The Death of Distance: how communications will change our lives.* London: The Orion Publishing Group Limited.

Cialdini, R. B., Borden, R. J., Thorne, A., Walker, M. R., Freeman, S., & Sloan, L. R. (1976). Basking in reflected glory: Three (football) field studies. *Journal of personality and social psychology*, *34*(3), 366.

Chompalov, I., Genuth, J., & Shrum, W. (2002). The organization of scientific collaborations. *Research Policy, 31* , 749-767.

Cohen, W., & Levinthal, D. (1990). Absorptive Capacity: A new Perspective on Learning and Innovation. *Administrative Science Quarterly, 35(1)* , 128-152.

Colombelli, A., & von Tunzelmann, N. (2011). The persistence of innovation and path dependence. *Handbook in the economic complexity of technological change* , 105-119.

Corley, E., Boardman, P. C., & Bozeman, B. (2006). Design and the management of multu-institutional research collaborations: Theoretical implications from two case studies. *Research Policy* , 975-993.

Corley, K., & Gioia, D. (2004). Identity Ambiguity and Change in the Wake of a Corporate Spin-Off. *Administrative Science Quarterly 49* , 173-208.

Christäller, W. (1933[1966]) Central places in Southern Germany. New Jersey: Prentice-Hal.

Crease, R. (1999). *Making Physics: A Biograhpy of Brookhaven National Laboratory.* Chicago: University Chicago Press.

Dahlander, L., & McFarland, D. A. (2013). Ties that last tie formation and persistence in research collaborations over time. *Administrative Science Quarterly*, *58*(1), 69-110.

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative Advantage as a Mechanism for Inequality: A review of Theoretical and empirical developments. *Annual review of Sociology* , 271-97.

Dosi, G. (1982). Technological paradigms and technological trajectories. *Research Policy, 11* , 147-162.

Erdös, R., & Rényi, A. (1959). On Random Graphs. *Publicationes Mathematicae Debrecen, 6* , 290-297.

Falk, A. and A. Ichino (2006). "Clean Evidence on Peer Effects." Journal of Labor Economics 24(1): 39-57.

Feldman, M. P., & Kogler, D. F. (2010). Stylized Facts in the Geography of Innovation. In B. H. Hall, & N. Rosenberg, *Handbook of the Economics of Innovation vol 1* (pp. 381-410). Amsterdam: North-Holland / Elsevier .

Galison, P., & Hevly, B. (1992). *Big Science. The Growth of Large Scale Research.* Stanford University Press.

Galunic, D. C., & Rodan, S. 1998. Resource recombinations in the firm: knowledge structures and the potential for schumpeterian innovation. *Strategic Management Journal*, 19(12): 1193-1201.

Glänzel, W., & Schubert, A. (2004). Analyzing Scientific Networks Through Co-Authorship. *Human Factors* .

Granovetter, M. (1985). Economic action and social structure: the problem of embeddedness. *American journal of sociology*, 481-510.

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology* , 1360-1380.

Groysberg, B., Lee, L. E., & Nanda, A. (2008). Can they take it with them? The portability of star knowledge workers' performance. *Management Science*, *54*(7), 1213-1230.

Groysberg, B., Lee, L. E., & Abrahams, R. (2010). What it takes to make star hires pay off. *MIT Sloan Management Review*, *51*(2), 57-61.

Henderson, J. (2007). Understanding Knowledge Spillovers. *Regional Science and Urban Economics, 37* , 740-760.

Henderson, R. M., & Clark, K. B. (1990). Architectural Innovation: The reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly, 35* , 9-30.

Hinds, P. J., Carley, K. M., Krackhardt, D., & Wholey, D. (2000). Choosing work group members: Balancing similarity, competence, and familiarity.*Organizational behavior and human decision processes*, *81*(2), 226-251.

Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. *Science, 322* , 1259-1262.

Katz, J., & Martin, B. (1997). What is Research Collaboration. *Research Policy 26* , 1-18.

Knorr-Cetina, K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. Boston: Harvard University Press.

Kilduff, M., & Krackhardt, D. (1994). Bringing the individual back in: A structural analysis of the internal market for reputation in organizations. *Academy of Management Journal*, *37*(1), 87-108.

Krugmann, P. (1996). The Self-Organizing Economy. Oxford, UK and Cambridge, MA: Blackwell.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Laband, D. N., & Tollison, R. D. (2002). Intellectual Collaboration. *The Journal of Political Economy, 108* , 632-662.

Langford, C., & Langford, M. (2000). The evolution of rules for access to megascience research environments viewed from Canadian experience. *Research Policy, 29(2)* , 169-179.

Laudel, G. (2002). What do we measure by co-authorships. *Research Evaluation* , 3-15.

Lauto, G., & Valentin, F. (2013). How Large Scale Research Facilities Connect to Global Research. *Annual Review of Research Policy* .

Link, A., & Scott, J. (2003). U.S. Science Parks: The Diffusion of an innovation and its effect on the academic missions of universities. *International Journal of Industrial Organization, 23* , 1323-1356.

Lomi, A., Snijders, T. A., Steglich, C. E., & Torló, V. J. (2011). Why are some more peer than others? Evidence from a longitudinal study of social networks and individual academic performance. *Social Science Research*, *40*(6), 1506-1520.

Lorenzen, M., & Andersen, K. V. (2009). Centrality and creativity: does Richard Florida's creative class offer new insights into urban hierarchy?. *Economic Geography*, *85*(4), 363-390.

McPherson, M., Smith-Lovin, L., & Cook, J. (2001). Birds of a Feather: Homophily in social networks. *Annual Review of Sociology, 27* , 415-444.

Merton, R. (1968). The Matthew Effect in Science. *Science, 159* , 56-63.

Merton, R. (1988). The Matthew effect In Science, ii: Cumulative Advantage and the symbolism of intellectual property. *Science, 79* , 606-623.

Merz, M., & Biniok, P. (2010). How technological platforms reconfigure science-industry relations: The case of micro- and nanotechnology. *Minerva, 48(2)* , 105-124.

Narin, F., Pinski, G., & Gee, H. H. (1976). Structure of the biomedical literature. *Journal of the American Society for Information Science*, *27*(1), 25–45.

Newman, M. (2002). Assortativity Mixing in Networks. *Physical Review Letters, 89* .

Newman, M. (2004). Coauthorship networks and patterns of scientific collaboration. *Proc. Natl. Acad. Sci. USA, 101* , 5200-5205.

Newman, M. (2001). The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA 98* , 404-409.

Padgett, J. F., & Powell, W. W. (2012). *The Emergence of Organizations and Markets.* New Jersey: Princeton University Press.

Phelps, C., Heidl, R., Wadhva, & Anu. (2012). Knowledge, Networks and Knowledge Networks: A Review and Research Agenda. *Journal of Management, 38* , 1115-1166.

Powell, W. (1990). Neither Market nor Hierarchy: Network Forms of Organization. *Research in Organizational Behaviour* , 295-336.

Powell, W., White, D., Koput, K., & Owen-Smith, J. (2005). Network Dynamics, and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. *American Journal of Sociology, 110* , 1132-1205.

Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology.*Administrative science quarterly*, 116-145.

Rawlings, C. M., & McFarland, D. A. (2011). Influence flows in the academy: Using affiliation networks to assess peer effects among researchers. *Social Science Research*, *40*(3), 1001-1017.

Robins, G., Pattison, P., & Woolcock, J. (2006). Small and Other Worlds: Global Network Structures from Local Processes. *American Journal of Sociology, 110* , 894-936.

Rosenberg, N. (1992). Scientific instrumentation and university research. *Research Policy, 21* , 381-390.

Salter, A. J., & Martin, B. R. (2001, vol 30 is 3). The economic benefits of publicly funded basic research: a critical review. *Research Policy* , 509-532.

Schwab, A. (2009). "Learning from broadcasts of project participation: selection of second-tier actors during project-venture formation in the movie industry, 1931-40." Strategic Organization **7**(2): 137-181.

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management science*, *51*(5), 756-770.

Simmel, Georg. 1955 (1922). Conflict and the Web of Group Affiliations. Translated by R. Bendix. New York: Free Press.

Skilton, P.F., 2008. Similarity, familiarity and access to elite work in Hollywood: Employer and employee characteristics in breakthrough employment. Human Relations 61, 1743-1773.

Smith-Doer, L., & Powell, W. (2010). Networks and Economic Life. In N. Smelser, & R. Swedberg, *The Handbook of Economic Sociology* (pp. 379-403). New York: Russel Sage Foundation & Princeton University Press.

Snijders, T., van de Bunt, G., & Steglich, C. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks, 32(1)* , 44-60.

Snijders, T. A. (2001). The Statistical Evaluation of Social Network Dynamics. *Sociological Methodology, 31* , 361-395.

Snijders, T. A., Steglich, C., & Schweinberger, M. (2007). Modeling the Co-evolution of Networks and Behavior. In K. V. Montfort, H. Oud, & A. Satorra, *Longitudinal Models in Behavioral and Related Sciences* (pp. 41-71). Lawrence Erlbaum.

Sorenson, O., D.M. Waguespack. 2006. Social structure and exchange: Self-confirming dynamics in Hollywood. Admi. Sci. Quart. 51(4) 560-589

Steglich, C., Snijders, T. A., & Pearson, M. (2010). DYNAMIC NETWORKS AND BEHAVIOR: SEPARATING SELECTION FROM INFLUENCE. *Sociological Methodology, 40* , 329-393.

Ter Wal, A. L., & Boschma, R. A. (2009). Applying social network analysis in economic geography: framing some key analytic issues. *The Annals of Regional Science*, *43*(3), 739-756.

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative science quarterly*, 35-67.

Wagner, C. (2008). *The New Invisivle College.* Brookings Institution Press.

Wagner, C., & Leyesdorff, L. (2005). Mapping the Network of Global Science: comparing international co-authorships from 1990-2000. *International Journal of Technology and Organization* , 185-208.

Wasserman, S., & Faust, K. (1994). *Social Netwrok Analysis: Methods and Applications.* New York: Cambridge University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393* , 440-442.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The Increasing Dominance of Teams in the Production of Knowledge. *Science, 316* , 1036-1039.

APPENDIX A

Table A1. List of variables.

Variable(s)	Description
Dependent network	**Co-authorship network - scientists have previously co-authored an article.**
Rate parameters	
Period effects	The three transition periods (2006-07, 2007-08, 2008-09)
NETWORK ESTABLISHMENT:	
Structural effects	
Triadic closure	The propensity to form ties with those whom one has had a prior indirect tie.
Preferential attachment	The preferential attachment effect (sqr(alter centrality))
Similar activity	Tendency for agents with high degrees to co-author with other agents with high degrees.
Performance effecs on structure	
Similar performance	Tendency to interact with those with similar # of citations.
Performance alter	Tendency of scientists to seek collaboration with other scientists performing at a higher level than themselves.
PERFORMANCE EFFECTS:	
Cumulative advantage	Captures the tendency of agents with high degrees to perform better at t+1
Similarity	Captures the tendency of agents to assimilate to their collaborators performance.
Controls	
Same organization	Whether scientists are from the same institution.
Same country	Whether scientists are from the same country.
Organization type	Tendency of scientist type to attract collaborators.
Performance x ego	Interaction effect showing the influence of performance on degree.
CHI x ego	Tendency for more basic scientists to establish collaborations.
CHI similarity	Tendency for scientists with similar CHI score to interact.

Table A2. Distribution of scientists pr. country

Country	N (authors)	Percentage
US	1290	59%
Germany	147	7%
Japan	141	6%
France	87	4%
China	81	4%
Rest	434	20%
TOTAL	2180	100%

Table A3.

Variable	Mean	Median	Stand. Dev.	Max	Min
Size of teams	6.70	6	5.80	97	1
Author publications	2.47	1	3.40	35	1
Performance	16.12	6	49.92	1052	6
CHI-level	3.17	3	0.90	1	4

Table A4. Network density indicators.¨

Observation time	1	2	3	4
Density	0.003	0.003	0.004	0.005
Average degree	9.301	11.483	12.605	16.784
Nmb of ties	31288	38628	42402	56462

Table A5. Network turnover frequency.

Periods	0 → 1	1 → 0	1 → 1	Jaccard
1 → 2	7360	20	31268	0.80
2 → 3	3782	8	38620	0.90
3 → 4	14074	14	42388	0.75

Table A6a. Results from stochastic actor based model.

Variable	Model 1 (controls only)		Model 2		Model 3	
	β	SE	β	SE	β	SE
Hypotheses 1a:						
Preferential Attachment			**-**		**-0.1785***	(0.05)
Hypotheses 2:						
Citations x alter			0.1369	(0.11)	0.1468	(0.10)
Citations x ego			**0.4030*****	(0.06)	**0.4625*****	(0.06)
Hypotheses 3a & 3b:						
Similar activity			**0.4852****	**(0.16)**		
Similar performance			**1.3202***	**(0.40)**	**1.5451***	(0.61)
Rate Parameter Controls:						
Period 1	1.4212***	(0.02)	1.2479***	(0.03)	1.3470***	(0.03)
Period 2	0.7013***	(0.01)	0.6758***	(0.01)	0.6959***	(0.01)
Period 3	1.9599***	(0.02)	1.9691***	(0.02)	1.9479***	(0.03)
Controls:						
Same country	0.8292***	(0.12)	0.7013***	(0.14)	0.7161***	(0.13)
Same organization	0.9311***	(0.13)	0.6250***	(0.14)	0.6006***	(0.13)
CHI x ego	0.9071**	(0.28)	1.4868***	(0.43)	1.2672***	(0.32)
CHI similarity	2.1204***	(0.27)	0.5066	(0.32)	0.4371	(0.32)
Triadic closure			0.5921***	(0.02)	0.5196***	(0.02)
Same researcher type controls	YES		YES		YES	YES

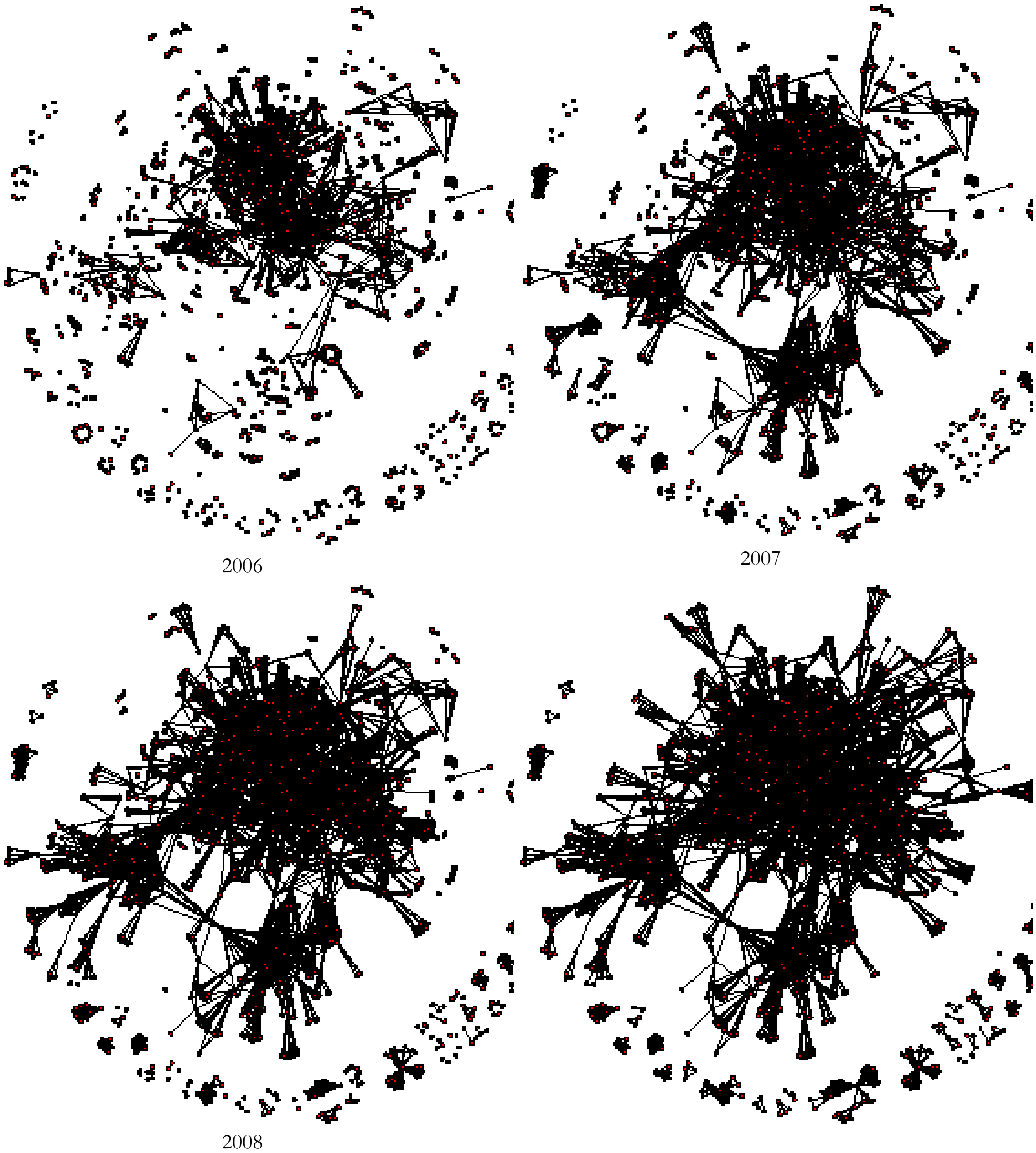
† p < 0.1; * p < 0.05; ** p < 0.01; *** p < 0.001

Table A6b. **Performance Dynamics**

	Model 2		Model 3	
Variable	**β**	**SE**	**β**	**SE**
Rate performance:				
Period 1	21.2305***	(11.67)	20.4705***	(11.66)
Period 2	5.2080***	(2.12)	5.3094***	(2.07)
Period 3	2.5941***	(0.11)	2.6357***	(0.08)
Linear Shape	-0.0892	(0.10)	-0.0410	(0.09)
Quadratic Shape	**0.1305***	(0.06)	**0.1216***	(0.04)
				(0.08)
Hypothesis 3c:				
Performance Similarity	**3.2799*****	**(0.08)**	**3.3797*****	(0.09)
Hypothesis 1b:				
Cumulative advantage	-0.0082	(0.01)	-0.0116	(0.01)

† p < 0.1; * p < 0.05; ** p < 0.01; *** p < 0.001

Table A7. Network visualizations (2006-09)



Layout: R-package Statnet and using the Fruchterman-Reingold algorithm.

Table A8. Correlation Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
rate collaboration																									
rate collaboration	0,54																								
rate collaboration	0,34	0,45																							
transitive ties	-0,16	-0,17	-0,16																						
preferential attachment	0,17	0,14	0,08	-0,57																					
same org	0,09	0,10	0,08	-0,19	0,02																				
same country	0,17	0,25	0,13	-0,05	0,07	-0,09																			
same resident	-0,14	-0,26	-0,20	0,08	-0,02	-0,10	-0,19																		
same university	0,08	0,02	0,01	-0,03	0,06	-0,16	0,06	0,07																	
same research lab	-0,05	0,02	-0,01	0,07	-0,06	-0,42	-0,05	0,04	0,07																
same secondmen	-0,01	0,05	0,03	0,03	-0,04	-0,23	0,02	-0,20	0,02	0,10															
same multiple	0,17	0,15	0,08	-0,10	0,17	-0,30	0,09	-0,06	0,06	0,17	-0,05														
citations alter	-0,04	-0,02	-0,05	0,12	-0,20	-0,04	-0,05	0,04	-0,05	-0,02	0,02	-0,02													
citations ego	0,61	0,71	0,44	-0,01	0,26	-0,11	0,31	-0,12	0,32	0,17	0,06	0,39	-0,09												
chi alter	-0,05	0,07	-0,02	0,20	-0,30	-0,01	0,05	0,00	-0,03	0,10	0,10	-0,10	0,03	-0,04											
chi ego	-0,38	-0,43	-0,34	0,12	0,01	-0,16	-0,17	0,27	0,19	0,08	-0,04	-0,04	-0,01	-0,42	0,05										
chi similarity	0,06	0,02	0,08	-0,36	0,21	0,08	0,06	-0,05	0,04	-0,05	-0,03	0,11	-0,13	0,04	-0,58	-0,11									
rate behavior	0,02	0,04	0,02	0,00	-0,02	-0,03	0,02	0,01	-0,05	-0,02	-0,04	0,05	0,02	0,00	0,00	0,00	0,02								
rate behavior	-0,10	0,05	-0,10	0,06	-0,03	-0,02	-0,13	0,09	0,10	0,10	0,16	-0,24	0,04	-0,12	0,45	0,13	-0,28	0,05							
rate behavior	0,11	0,09	0,07	0,03	-0,02	-0,05	0,04	0,00	0,02	0,07	0,09	0,04	-0,02	0,12	0,08	-0,06	-0,09	0,01	0,05						
linear shape	0,05	0,04	0,02	0,00	0,01	0,06	0,04	0,05	-0,01	-0,06	0,00	0,04	-0,01	0,06	0,02	0,01	-0,02	0,10	-0,08	-0,19					
quadratic shape	0,15	0,07	0,01	-0,08	0,11	-0,07	-0,08	0,07	0,02	-0,09	0,02	0,12	-0,04	0,05	0,08	0,01	-0,05	0,30	0,22	0,02	0,13				
performance average similarity	-0,04	-0,12	-0,14	0,06	-0,02	-0,08	-0,06	0,13	0,05	-0,09	0,04	-0,09	0,01	-0,11	0,04	0,12	-0,01	0,15	0,16	0,02	0,22	0,47			
Performance from Outdegree	-0,08	-0,05	-0,04	0,04	-0,05	-0,05	-0,07	0,00	0,03	0,10	0,01	-0,08	0,05	-0,08	0,05	0,04	-0,02	-0,15	0,18	0,19	-0,89	-0,21	-0,19		
Performance from CHI	-0,03	-0,02	0,06	0,03	-0,06	0,02	0,07	-0,08	-0,10	-0,05	0,03	-0,02	-0,02	-0,02	-0,09	-0,07	0,02	-0,08	-0,21	0,03	0,15	-0,14	0,00	-0,30	