



Paper to be presented at the DRUID Academy conference in Rebild, Aalborg, Denmark on January

15-17, 2014

## **Big Egos in Big Science**

**Jacob Emil Jeppesen**

Copenhagen Business School  
Department of Innovation & 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**

Copenhagen Business School  
Department of Innovation and Organizational Economics  
fv.ino@cbs.dk

### **Abstract**

In this paper we investigate the micro-mechanisms governing the structural evolution of a scientific collaboration. Empirical evidence indicates that we have transcended into a new paradigm with a new modus operandi where scientific discovery are not lead by so called lone ?stars?, or big egos, but instead by a group of people, from a multitude of institutions, having a diverse knowledge set and capable of operating more and more complex instrumentation Using a dataset consisting of full bibliometric coverage from a Large Scale Research Facility, we utilize a stochastic actor oriented model to estimate both the structural and performance effects of selection, as well as the behavioral of crossing organizational boundaries. Preliminary results suggest that the selection of collaborators still is skewed, and identify a large assortativity effect, as well as a tendency to interact with both authors with similar citations.

# Big Egos in Big Science\*

\*Preliminary work. Do not quote or circulate without author's permission

**Jacob Jeppesen<sup>\*1</sup>, Kristina Vaarst Andersen<sup>1</sup>, Giancarlo Lauto<sup>2</sup> & Finn Valentin<sup>1</sup>**

<sup>1</sup>Copenhagen Business School, Department of Innovation and Organizational Economics, Kilevej 14A, 2000 Frederiksberg, Denmark.

<sup>2</sup>University of Udine, Dipartimento di Scienze Economiche e Statistiche, Via Tomadini 30/A, Udine, Italy.

\*Corresponding author. Email: [jj.ino@cbs.dk](mailto:jj.ino@cbs.dk)

## **Abstract**

In this paper we investigate the micro-mechanisms governing the structural evolution of a scientific collaboration. Empirical evidence indicates that we have transcended into a new paradigm with a new modus operandi where scientific discovery are not lead by so called lone 'stars', or big egos, but instead by a group of people, from a multitude of institutions, having a diverse knowledge set and capable of operating more and more complex instrumentation

Using a dataset consisting of full bibliometric coverage from a Large Scale Research Facility, we utilize a stochastic actor oriented model to estimate both the structural and performance effects of selection, as well as the behavioral of crossing organizational boundaries. Preliminary results suggest that the selection of collaborators still is skewed, and identify a large assortativity effect, as well as a tendency to interact with both authors with similar citations.

## Introduction

The notion of the lone genius creating artistic masterpieces, inventions or scientific breakthroughs in solitude has long prevailed, though research continues to demonstrate that creativity of all sorts predominantly is grounded in collaboration. Through collaboration individuals reach the scale and scope necessary to achieve results impossible for any individual. Collaboration allows teams to join resources and tackle problems too big for each of them, and collaboration further allows individuals with different expertise to combine their respective abilities and tackle problems too complex for each of them.

Analyses show an increasing tendency for collaborative research, but further there is evidence of collaborative projects having more impact than individual/pair lead research, and of boundary spanning collaborations to have the highest impact. Most papers and patents are now developed in collaborative teams as a result of the increasing tendency for collaboration (Wutchy et al., 2007). Collaboration facilitates division of work and the pooling of intellectual expertise, it permits the accomplishment of projects that could not be realized by a lone scientist (Katz & Martin, 1997), and increase the number of studies and hence the chances and number of published studies for the individual researcher (Barnett et al., 1988). This is evident in all fields of research, but especially for interdisciplinary studies or research involving specific instrumentation (Chombalov et al., 2002). Collaboration often transcends organizational and national borders, but reinforces stratification between elite institutions and the rest (Jones et al., 2008). Even in the collaborative world of big science we find big egos.

The tendency for collaborative research has been mapped and evidence presented that collaborative research especially when combining mainstream knowledge with atypical knowledge increase impact, but the micro-foundations for research collaboration remain unexplored. Knowing not only the results of collaborative effort, but also the individual level mechanisms through which these collaborations came into being is pivotal for understanding collaborative research and the causes of the evident stratification in collaboration. In this paper we analyze the microfoundations of collaborative research. 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 collaboration and find that various forms of homophily in are main drivers of establishing team work and consequently the creation an inter-

connected elite of high performing researchers found in Jones et al. (2008). 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 an actor's network position, and local network structure to the network evolution and performance of individual scientists.

To analyze the micro-foundations of research collaboration, we turn to a setting purposefully set up to foster collaboration across organizational and national borders: Large Scale Research Facilities. We consider LSRF a critical case for testing homophily tendencies in research collaboration as they are designed to facilitate open 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, with the left tail of research performance distribution omitted, 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/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 effects of homophily and the rules of cumulative advantage in establishment of ties between collaborating researchers. If big egos exist here, we should expect them to be predominant throughout all scientific fields and settings.

## Theory

Recently, a literature stream has started to emerge with focus on using the dynamics of network emergence and evolution as explaining collaboration and innovation (Ahuja, Soda, & Zaheer, 2012). More recent research addressing the macro-dynamics of networks to understand how e.g. organizational fields evolve (Powell et al. 2005), and from a more micro-dynamic perspective, the extent to which knowledge flows are geographically mediated (Azouley et.al., 2011). But looking at the micro-mechanisms driving the network evolution of scientific collaboration has yet to be done. As actors

usually are not able to cast their gaze across the entire network, on the basis of their localized view they form ties and make decision based upon the intersection with those that are socially proximate (Robins et.al., 2005).

### **The notion of the lone genius**

Both the Kuhnian and Dorian view has been studied from numerous perspectives and from multiple levels, and over the years a large pool of knowledge has been established. One of the more persistent notions is that of the lone genius as driver of scientific discovery. This has long been tradition in both the history and philosophy of science, and has a tendency of equating great ideas to sole scientists, as seen in the form of e.g. the Nash Equilibrium, Arrow's theorem of impossibility, Newtonian mechanics and Einstein's theory of relativity. With the increased professionalization of science post-World War II, a new term for describing the focus on research collaboration emerged: *big science*. The term Big Science describes a series of changes in the scientific community that occurred during and after World War II. The making of science shifted from individual or small group efforts, "Small Science", to relying heavily on large scale research projects, mostly characterized by their extremely high monetary costs and large increase in the number of collaborating, often international, partners (Barabási, 2002; Newman M. , 2001; Wagner & Leydesdorff, 2005).

Following, more and more empirical evidence indicates that we have transcended into a new paradigm, where scientific discovery is not lead by so called lone 'stars', or big egos, but instead by a group of people, from a multitude of institutions, having a diverse knowledge set and capable of operating more and more complex instrumentation (Katz & Martin, 1997; Newman, 2002, 2004). Thus the change in the mechanisms governing the production of scientific knowledge are not only found at these large scale collaboratives, but has also spilled over into regular science (Wuchty et.al., 2007).

With the locus of innovation thus located in collaborative networks, this pends the two critical questions: How and why modern scientific collaborations evolve and take the forms they do and what separates a highly innovative scientist from a less so?

## Research challenging the notion of the lone genius

Extensive literature has dealt with this issue, both from a scientific collaboration perspective and with other innovative units as focus. Under the assumption that the premises for collaboration are similar, empirical findings show e.g. that diversity of knowledge can facilitate innovation through recombination (Henderson & Clark, 1990), that structural positions and increasing the number of co-authors result in increased scientific production and impact (Wuchty et.al., 2007; Abbasi et.al., 2011), that the number of organizational boundaries crossed are negatively related to innovation unless the collaborators taken together spans otherwise distant units (Bercovitz & Feldman, 2011). Despite this accumulation of knowledge on the collaborative process, there is still little understanding of the role of structural dynamics (like transitivity and preferential attachment) in shaping the structure and formation of collaboration and knowledge creation (Phelps et.al., 2012; Ahuja, Soda & Zaheer, 2012). The structural understanding has thus far been dominated by an intent on reproducing the topological form of real world networks (e.g. Erdős & Rényi, 1959; Watts & Strogatz, 1998; Barabási et.al., 2002). This form has largely ignored an extensive tradition in both the literature from sociology, psychology and economics regarding the behavior and characteristics of individuals. In this paper we employ an alternative approach allowing us to model both the structural, nodal and behavioral characteristics. Thus we are able to represent network and behavior change as the result of dynamics being driven by different tendencies and especially structurally based micro-mechanisms.

## Collaboration in research

The earliest collaborative paper appearing in a scientific journal ever to be recorded was a paper in the *Philosophical Transactions of the Royal Society*, by Hooke, Oldenburg, Cassini & Boyle published in 1655 (Beaver & Rosen, 1978). At that time collaboration were driven by totally different rules than we see today. Spatial constraints, especially geographically but also the mere structure and availability of education greatly inhibited the possibilities for collaboration. Fastening forward, an especially large change in these premises can be seen after WWII. An increasing professionalization of science, and the increase in overall funding, has resulted in the absorptive capacity (Wagner & Leyesdorff, 2005) and interconnectedness of the scientific community as a whole to increase enormously.

This has resulted in not only the whole community of science being more interconnected, but also the sciences themselves becoming more and more inter-dependent, and thereby reliant on the focal scientist's ability to connect and collaborate (Wagner &

Leydesdorff, 2005; Newman, 2001, 2004). On the dark side of this, empirical findings also show that despite the rising frequency of inter-university collaboration, this is not driven as a result of increasing equality or reduced coordination costs. Instead it is mostly governed by an intensification of social stratification with a concentration of the production of scientific knowledge in few high-prestige centers of high impact science (Jones et.al. 2008). While this has been shown from a university perspective, little research has been done from an evolutionary micro-perspective, showing the governing motivations for single researchers to collaborate. Following, a question in the need of further treatment is why scientists increasingly choose to collaborate?

### Reasons to collaborate

In their literature review of collaborative research, Katz & Martin (1997) note that in essence collaborative research has become the model per se in many fields of science due to specific benefits: **First** that it facilitates division of work and the pooling of intellectual expertise. **Second** they note that collaborating permits the accomplishment of projects that could not be realized by a lone scientist. This can be seen especially in interdisciplinary studies or research involving specific instrumentation (Chombalov et.al., 2002). Collaboration increases the number of studies that can be undertaken and therefore, the probability that an author's work will be accepted for publication in a journal (Barnett et al., 1988). The empirical findings in this fields also points in the direction that co-authored papers present a higher quality than those which are single-authored (Laband and Tollison, 2000), which leads to a higher impact (Wutchy et.al., 2007; Katz and Martin, 1997).

While this may suggest an existence of spatial constraints on collaboration, favoring face-to-face contact and the enforcing of the "30-feet collaboration rule", recent empirical evidence has given more sound to the "death of distance" point (Cairncross, 1997), showing a remarkable rise in the inter-university collaboration since especially 1975 (Jones et.al., 2008). In the same study the authors also find that research conducted in-between universities are more likely to increase average citation score, suggesting an increased performance when research is conducted in teams spanning multiple institutions.

## Hypotheses

### Revisiting the Matthew Effect

One of the most established facts in network theory is the concept of cumulative advantage. Prior research has shown a highly skewed distribution of productivity among scientists, resulting in an effect where highly productive researchers maintain or increase their productivity while scientist who produces very little produce even less later on (Allison & Stewart, 1984). Even though originally developed as a means to explain advancement of scientists by Robert Merton in 1968, the notion of the Matthew Effect, i.e. the rich gets richer while the poorer gets poorer, has been shown to have general applicability as a mechanism for inequality across many temporal processes (DiPrete & Eirich, 2006). 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 establishment of centralized research centers, as seen in the case of Large Scale Research Facilities, the epitome of the new paradigm in science, has often been instigated to negate this unequal division, due to its formalization of collaboration (Lauto & Valentin, 2013).

On the nodal level, preferential attachment increases researchers' tendency to seek out highly central new collaboration partners. When new ties are formed, they tend to be directed towards researchers who already have many collaboration partners and are central to the social structure. Researchers who are already central to the network will have many opportunities for collaboration, and potentially also opportunities to pick the most promising collaboration partners. Based on this Rich-gets-richer mechanism, we propose:

**H1: researchers who are central in the collaboration network experience increased probability of forming new ties (preferential attachment).**

### Proximity and distance

Proximity and distance between interaction partners have been shown to affect the probability of tie formation and outcome, in that proximity increases the probability of tie formation, but the ties established across distance and boundaries tend to result in higher performance. Geographical proximity increases probability and frequency of random encounters potentially resulting in collaboration. Organizational proximity share this feature and further serves as a framework for commonality of norms and incentives.



Finally, cognitive proximity decrease costs of interaction and increase efficiency (Lauto & Valentin, 2013; Bercovitz & Feldman, 2011). Based on this we develop three hypotheses:

**H2a: Common organizational affiliation increases the probability of tie formation.**

**H2b: Common institutional environment increases the probability of tie formation.**

**H2c: Cognitive proximity of research fields increases the probability of tie formation<sup>1</sup>**

### **Birds of a feather..**

Recently empirical findings in social networks show the existence of a much more skewing effect, namely that of the “rich club effect”, meaning that prominent nodes direct their ties towards each other (Newman, 2002). Finding significance for the existence of this in the real world network of scientific collaboration, especially in a world instigated to negate this, would give an indication of a tendency both an increasing inequality – i.e. scientists that does prefer to collaborate, even though their merits can be equally high, are increasingly eased out of the network.

Thus we hypothesize:

**H3a: Highly active researchers tend to form new collaborations with other highly active researchers.**

To tease out to which extent this rich club effect is either a phenomenon based on social status or performance we will also employ the number of citations, and test for whether the network are dominated by performance homophily:

**H3b: High impact researchers tend to form new collaborations with other high impact researchers.**

### **Empirical 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, Tennessee. Established in 1943, the facility is a multidisciplinary center financed by the U.S.

---

<sup>1</sup> The model included in this paper does not test for this hypotheses.

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 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. 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.

### **Big Science and Large Scale Research Facilities**

When the 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 the different instrumentation and in the 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 in science earlier mentioned.

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). 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, as shown, radical potential. Research in multi-institutional collaboration in the natural sciences has been primarily dominated by historians, sociologists and anthropologists, focusing on in particular on high-energy particle physics (Chombalov 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 even been described as an example of the new model for collaboration in science (Knorr Cetina, 1999). This notion has since been challenged in (Chombalov et. al., 2002) in where it is shown that this mode of organizing in multi-institutional research projects is the exception of the rule, and largely found in the HEPP community.

## 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. Collaboration on articles creates a social network, the study of which allows us to understand some of the characteristics of a particular discipline or research site, to identify the invisible colleges (Wagner, 2008) and social groups that exist in all scientific fields. Studies in this have shown the potential of using social network analysis in opening up an interesting line of investigation in this respect (Barabási et al., 2002; Newman, 2001). Yet, the research specifically on structural integration, social homophily and how ability affects this, has been hampered by a lack of longitudinal analysis, with analysis up till now mainly consisting of snapshots. Not having a longitudinal perspective will greatly reduce the ability to causally infer the direction of selection and influence (Borgatti & Halgin, 2011). Indeed separating these mechanisms is central to addressing the issue of endogeneity in network papers (Steglich et.al. 2010). But to the best of our knowledge, no studies have combined a longitudinal network framework studying the evolution of scientific collaborations, incorporating both structural and behavioral effects. Thus the network effects of e.g. transitivity and preferential attachment will skew the results when not properly controlled for. 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), 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 ORNL/NSD. Since 2006, all peer-reviewed publications based on research utilizing ORNL/NSD data and resources, or conducted by staff affiliated with ORNL/NSD are publicly listed on the directorate's website. We refer to these publications as ORNL/NSD-based research. We retrieved full bibliometric records from ISI-Web of Science of the publications produced in the period from 2006 to 2009. Due to the calculative complexity of the simulation models, it was necessary for the trial run used in this paper to limit the total amount of publications. The criteria for selection were thus set as a) at least 4 citations b) each author should at least figure twice the first year of appearance and c) each author should at least be present in two time periods. This left us with a total of 108 distinct authors and 439 publications.

## Method: Modeling Dynamic Networks

The fundamental network consists of only two basic elements – the nodes and ties between these nodes (Wasserman & Faust, 1994). The nodes represent some actor and a tie between two actors suggests the existence of a flow or bond, in our instance a co-authorship tie.

To model the temporal dynamics of networks at the LSRE, I apply a stochastic actor-based approach. Here, the evolution of social networks, in terms of tie establishment and termination between the different actors, is 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 to capture endogenous effects, which are of high importance when explaining the evolution of social networks, as mentioned earlier. Even though this is a powerful analytic tool, some fundamental underlying assumptions have to be met (see also Snijders et al., 2010):

First, the network under analysis evolves as a stochastic process driven by the actors, which have control over their ties. This fundamentally implies that ties are directed, hence sent by one actor and received by another, where the former controls the tie establishment. Here the methodology of using one-mode projections of a two-mode network (resulting in undirected networks) would basically violate these assumptions. But as proposed in (Snijders, 2010) this can be controlled for through the choice of the pairwise conjunctive model, where a pair of actors is chosen and reconsiders whether a tie will exist between them. The tie will exist if both agree, and it will not exist if at least one

does not choose it. We lose, of course, information on who made the initial contact, and the interpretation of the results should reflect this potential of selection bias.

Second, tie changes are assumed to be a gradual process, taken in the form of a series of mini-steps – hence modeled in continuous time. This is usually valid for persistent relationships such as friendship, trust, strategic alliances et cetera. In contrast, relationships based on event data, such as phone calls, e-mails or, as in this case, co-authored publications are a non-replicable event, and hence in general cannot be interpreted as enduring. Nevertheless, co-authorship is in this instance seen as a proxy for more enduring relationships – potentially both friendship and/or professional. In order to accommodate this, a co-authorship is in this data seen as enduring – i.e. having established a tie at time  $t$  means the tie is persistent at time  $t+1$ . This will tend to overestimate the number of established ties as no dissolutions are allowed, and thus the results will tend to overestimate the effect of behavior change.

### Separating selection from influence

Stochastic actor-based networks basically consist of some a rate function, controlling the changes in the network, an objective function consisting of a set of individual parameters  $\beta_k$  which determine how likely it is for an actor  $i$  to change their own ego-network in a particular way, and a behavior function. The decisions are modeled as the outcome of changes made by actors in a series of micro-steps. In the decision process,  $i$  has the opportunity to choose between some set  $\mathbf{C}$ , containing all possible ties with other network actors to remain either unchanged or change from being absent  $(0, x_a)$  to present  $(0, x_b)$ , and vice versa. Almost at the same time they have the opportunity to change their behavior, and either increments or decrements his or hers score on the behavioral variable.

Consequently, the probability of the overall network to change to some new state  $X$  or some new behavior  $Z$  is given by the formula:

$$(1) \quad \Pr(x, z \rightarrow x', z) = \frac{\exp(f_i^x(x', z))}{\sum_{x''} \exp(f_i^x(x'', z))}$$

$$(2) \quad \Pr(x, z \rightarrow x, z') = \frac{\exp(f_i^z(x, z'))}{\sum_{z''} \exp(f_i^z(x, z''))}$$

It basically resembles a multinomial logistic regression, modeling the probability that an actor chooses a specific (categorical) new network configuration  $X$ , or new behavior  $Z$ , as proportional to the exponential transformation of the resulting network's objective or behavior function. This model for tie probabilities was also used by Powell et.al., 2005.

### **Dependent Variables**

As the modeling in the SIENA-methodology employed in this paper are given by a) the network dynamics and b) the behavioral dynamics, these are subjects to change almost simultaneously, and thus making it possible to control for the structural dynamics of networks, e.g. transitivity and preferential attachment. Thus I model the dependent variable as mentioned in equation (1) and the behavioral dynamics are modeled as mentioned in equation (2) with a transformation of the citations variable into an ordinal, on a scale of 1-5<sup>2</sup>.

### **Key Variables**

The individual parameters can be divided into three categories: (i.) Network base effects, referring to the actors 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 behavioral. In the following I discuss in detail the main effects included in the model.

Additionally to these structural variables, another set of degree related measures are of particular interest against the background of this study. CENTRALITY (ALTER) represents the tendency of actors to form ties to alters already receiving a high amount of in-degrees, hence popular ones. A positive alter popularity implies a self-reinforcing mechanism that over time leads to increasing dispersion of the degree distribution of the networks. It can be interpreted as the impersonation of the Matthew Effect or preferential attachment in network structuralism. This effect has been shown numerous times to drive co-authorship networks and acts as a structural control for teasing out the assortativity effect.

---

<sup>2</sup> This effect has yet to be included in this working paper.

DEGREE ASSORTATIVITY refers to the preference of actors to form ties with alters based on their own as well as the alters degree. Because this is an un-directed network only the out-out degree measurement can be used. The out-out combination represent a measurement for homophily and social stratification in the network pattern, and in the case of the LSRF that active social scientists (in the form of co-authorships) team up with scientists with similar social activity.

TRANSITIVITY is another 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.

CITATION SIMILARITY 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).

SAME ORGANIZATION is a binary variable indicating whether or not the scientists stem from the same institutions.

SAME ORGANIZATION TYPE is a binary variable 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.

### **Base effects**

The baseline effect is given with the DENSITY, or sometimes called degree, representing the general tendency to form ties at all. It can be interpreted as the benefits and costs between an arbitrary tie. Arbitrary means in this context a tie with an actor embodying no characteristics making him/her particularly attractive.

**For mathematical notation of effects see the RSiena Manual (Snijders et.al., 2013).**

### **Controls**

We include a number of controls to account for the possibility of spuriousness, alternative sources of influence and selection. To control for the propinquity of researchers choosing to collaborate due to different notions of homophily we control for whether they are the same SCIENTIST TYPE (measured as the average CHI-score). We also control for the number of citations (ln) received by researchers and for whether they are STAR SCIENTISTS meaning being in the 95<sup>th</sup> percentile.

Table 1 depicts the complete list of variables included, and to be included, in the model:

<< INSERT TABLE 1 HERE >>



## Findings

Table 2 shows the results of a multinomial logistic regression based on the stochastic actor-based approach, with the behavioral function shown in table 3.

<< INSERT TABLE 2 HERE >>

All parameter estimations are based on 1,000 simulation runs. Convergence of the approximation algorithm is excellent for all the variables of the different models. It indicates whether the deviation of the simulated structures compared to the observed structures is acceptable (t-values  $< 0.1$ ), and can be used to evaluate the goodness of fit of the different models. 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. Odds ratio can be computed as the exponentiated form of the coefficients of each predictor. The estimated effect of citation similarity shows that similarity in amount of citations increases probability of researchers forming ties. We further see that transitivity is greatly significant meaning a high degree of clustering is present in our network. Contrary to our hypothesis, the measurement for preferential attachment comes out insignificant, meaning that the forming of network ties is not significantly guided by the establishment of ties to highly active researchers. At the same time though, the significance of the assortativity effects gives indication of the existence of a highly elitist effect where the scientists collaborating with the most predominantly chooses to collaborate with others of similar social visibility. We also find significance of the different proximity measures based on being from the same organization or being from the same institutional type, indicating that organizational proximity of its various forms are a significant driver of tie formation.

## Discussion

The literature generally argues for a tendency of network actors to form network ties with others primarily based on some kind of homophily – either based on general characteristics (e.g. gender, age) (McPherson, et.al., 2001) or network structure (Burt, 1982). In this network we see that scientists with similar citations are more likely to

collaborate, indicating a more hierarchical than competitive environment. This can be seen in the context of the specificities of the facility i.e. that scientists working employed as experts on specific instruments usually are co-authored, making them the hubs of the Big Science world.

The high degree of clustering in our network aligns with existing theory on small worlds particular persistent in networks based on scientific collaboration (Watts and Strogatz, 1998).

The insignificance of preferential attachment has to be seen together with the significance of assortativity (the tendency for highly active actors to form ties with other highly active) which suggests the existence of a “rich boys effect”, and that the creation of ties at the Spallation Neutron Source are more governed by social visibility, where scientist with large collaborative capacity choose to collaborate with other scientists with similar capacity. This creates a highly clustered and closed network, which is surprising given the paradigmatic change of scientific collaboration predicted in the theory building section. Potentially confounding factors such as discipline could serve as further inquiry into this relationship as figure 1 shows a high a degree of local clustering (transitivity) as well.

At facilities like SNS it is highly likely that the disciplines have a hard time finding convergence, i.e. the physicist focusing on crystallography or X-rays can have a hard time finding collaboration with micro-biologists. Taken together, the scientific collaboration network established at the Spallation Neutron Source seems to be governed much by the notion of homophily and a form of hierarchy based on performance. This should not necessarily be taken as a negative effect, as 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 a form of core group, making dissemination of knowledge happen faster but at the cost of the size of the giant component (Matthew, 2002). But as 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 finding 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 with scientists of similar performance could even mean less inter-disciplinarity. But future analysis including scientific discipline is needed in directing this.

Taken together our results suggest that connecting is not only a function of performance, but more of a social process governing the evolution of the entire network, much in line with the findings of Jones et.al., 2008.

## 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 showed 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. At the same time we concluded the need for a longitudinal perspective on if we are to say anything meaningful of the influence of networks on scientific collaboration. We proposed the hypotheses that science was 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”-effect.

We investigated 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 conducted the analysis based on a selection of the full bibliometric recording<sup>3</sup>, going from 2006-09, of publications affiliated with the Spallation Neutron Source and the High Flux Isotope Reactor.

To test our hypotheses we employed a stochastic actor based network analysis to separate the selection and network dynamics from the behavior of external collaboration. By doing so we were able to analyze the cumulative and self-reinforcing effects of network dynamics. We find that we indeed see a network dominated not by the Matthew Effect, but even more both by assortativity and citation similarity, indicating both a highly unequal distribution of collaboration but also one dominated by some form of scientific hierarchy. We also found that the number of external collaborators is positively related to the formation of a tie.

---

<sup>3</sup> In this paper, only a subset of the nodes is used. See appendix 2 for description of selection.

### **The future directions of the paper**

As mentioned in a series of footnotes in this paper, we still need to do substantial work on the SIENA-model. Besides running the model on the full dataset, we also need to include the behavioral function of performance. The current model are based upon citations of publications in the same year as they appear. We are working on including a lagged citation variable instead, to truly model the effect of performance on network evolution, as scientists must be thought of as not being truly able to assess the impact papers before publication.

Besides this, we also tend to include an analysis of the effect of cognitive distances, both in the analysis of network evolution and performance, based upon keyword analysis of papers published by the scientists at the SNS. Thus we are working both on the scripting of this in RSIENA, but we are also working on collecting full bibliometric data, not only on publications affiliated with SNS, but full bibliometric data on the authors affiliated with SNS.

We also tend to code demographic data into the model, like gender and tenure based upon the full publication record of scientists.

## References

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

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

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 multi-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.

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

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.

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

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.

Glänzel, W., & Schubert, A. (2004). ANALYZING SCIENTIFIC NETWORKS THROUGH CO-AUTHORSHIP. *Human Factors* .

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

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.

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.

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.

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. *isis*, 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.

- 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.
- 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.
- 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.
- Sneijders, 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.



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

Wagner, C. (2008). *The New Invisible 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 Network 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 1

Table 1. Exhaustive list of variables. Variables marked with a \* are not incorporated in the model yet.

Variable(s)	Description
Dependent Network	Co-authorship network - Scientists have previously co-authored an article.
<b>Rate Parameters</b>	
Period Effects	The three transition periods (2006-07, 2007-08, 2008-09)
<b>Structural Dynamics</b>	
Density	The out-degree effect, basically the propensity of the network actors to form ties.
Transitivity	The propensity to form ties with those whom one has had a prior indirect tie.
Alter centrality	The preferential attachment effect ( $\sqrt{\text{alter centrality}}$ )
Assortativity effect	Reflects the tendencies for actors with high degrees to co-author with other actors with high degrees.
*Betweenness	Represents brokerage, i.e. the tendency for actors to position themselves between not directly connected others.
*Knowledge overlap	The existence of a direct reference between ego and alter.
*Knowledge similarity	Tendency to interact with those with similar quantities of keywords in knowledge area $k$ .
Scientist focus	CHI based on publications at the facility (Avg.>2 = applied scientist else basic.)
*Scientist type	1= Resident, 2=Secondment, 3= Single, 4=Multiple
Same organization type	1= Facility, 2=University, 3=Research Lab, 4=Business
*External collaboration	Count of # of external collaborators
<b>Performance</b>	
Similar citations	Tendency to interact with those with similar # of citations (fractional count).
*Research Productivity	Number of publications for author $i$ divided by total number of co-authors.
<b>Other effects</b>	
Same organization	Whether the two scientists are from the same institution.
Scientist type similarity	Tendency to collaborate with scientists of the same type.
Similar productivity	Scientists have similar levels of productivity in publications.
Star Scientist	Whether the scientists are in the 95 <sup>th</sup> percentile of high performers
*Scientist focus similarity	Tendency to collaborate with scientist with same focus.
<b>*Performance (behavior function)</b>	
*Period Effects	The four transition periods (2007-08, 2008-09, 2009-10, 2010-11)
*Linear Shape	The basic drive toward external collaboration.
*Quadratic Shape	The effect on the behavior on itself, as either self-limiting or self-reinforcing.
*Knowledge similarity	Tendency to interact with those with similar quantities of keywords in knowledge area $k$
*Scientist Focus	Avg. CHI based on publications at the facility (>2 = applied scientist else basic.)

Table 2. Results from SIENA Model predicting tie formation

Variable	B	SE
<b><i>Hypotheses 1:</i></b>		
Centrality (alter)	0.4735	(0.39)
<b><i>Hypotheses 2a, 2b &amp; 2c:</i></b>		
Same Organization	<b>1.2411*</b>	<b>(0.76)</b>
Same Organization type	<b>0.1325*</b>	<b>(0.06)</b>
Cognitive distance	-	-
<b><i>Hypotheses 3a &amp; 3b:</i></b>		
Assortativity	<b>1.543**</b>	<b>(0.28)</b>
Similar citations	<b>2.421**</b>	<b>(0.67)</b>
<b><i>Rate Parameter Controls:</i></b>		
Period 1	<b>1.2814**</b>	<b>(0.10)</b>
Period 2	<b>2.6304**</b>	<b>(0.16)</b>
Period 3	<b>0.7100**</b>	<b>(0.08)</b>
<b><i>Structural Controls:</i></b>		
Density (Degree)	<b>-2.9748**</b>	<b>(0.56)</b>
Transitivity	<b>0.9219**</b>	<b>(0.20)</b>
<b><i>Monadic and dyadic controls:</i></b>		
Star scientist	<b>3.1221**</b>	<b>(0.13)</b>
Same Scientist focus	-0.12	(0.1)
# of Citations (ln)	<b>2.1132*</b>	<b>(0.31)</b>

N=108

\*  $p < 0.05$ .

\*\*  $p < 0.01$

Rate parameters above zero are always significant

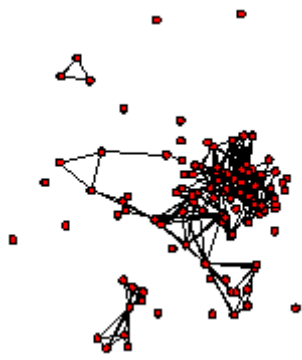
Table 3. Network density indicators

Observation time	1	2	3	4
Density	0.037	0.061	0.1110	0.124
Average degree	3.907	6.467	11.720	13.140

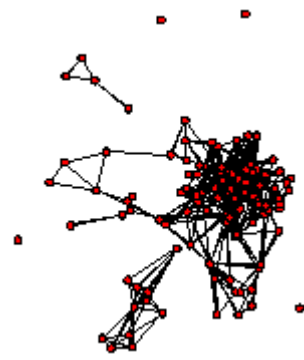
Table 4. Network turnover frequency

Periods	0 → 1	1 → 0	1 → 1	Jaccard
1 → 2	137	0	209	0.604
2 → 3	281	0	346	0.552
3 → 4	76	0	627	0.892

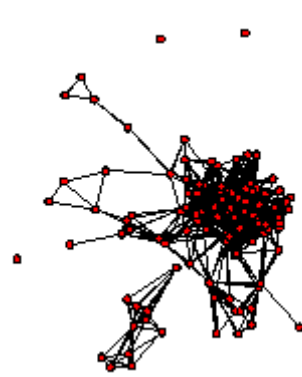
Table 6. Network visualizations (2006-09)



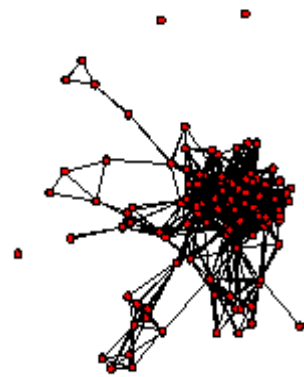
2006



2007



2008



2009

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

## Appendix 2

The current implementation of the SIENA-model is based upon a series of criteria for selecting the population:

1. Scientists have to publish at least two articles the first year of appearance.
2. Scientists have to appear with published articles at least for two years.
3. Scientists must be in the giant component of the final year.
4. Scientists must have achieved at least 4 citations for the final year.

The reason for doing this at this point is first and foremost to reduce the amount of nodes. Due to the computational complexity of the algorithms in the SIENA model the full model would take about 7 days, thus based on time constraints we chose this methodology.