# Competitive project funding and dynamic complex networks: evidence from Projects of National Interest (PRIN)

Antonio Zinilli[1,2]

**Abstract** This paper aims to study the collaboration among researchers in a specific Italian program funding, the Projects of National Interest (PRIN), which supports the academic research. The paper uses two approaches to study the dynamic complex networks: first it identifies the observed distribution of links among researchers in the four areas of interest (chemistry, physics, economics and sociology) through distribution models, then it uses a stochastic model to understand how the links change over time. The analysis is based on large and unique dataset on 4322 researchers from 98 universities and research institutes that have been selected for PRIN allocation from 2000 to 2011. The originality of this work is that we have studied a competitive funding schemes through dynamic network analysis techniques.

**Keywords** Competitive project funding · Dynamic complex systems · Power-law distribution · Preferential attachment · Stochastic actor oriented model

## Background and aim of the paper

Although the study of complex networks can count on a long tradition in various fields of knowledge, in the last few years, after the introduction of models for power-law (Barabàsi and Albert 1999) and scale-free networks (Watts and Strogatz 1998), complex networks have become increasingly important in many fields of science including social sciences, biology, chemistry, computer science etc. Complex networks are used to describe different real world systems (Albert and Barabási 2002). By this, we mean complex network systems with a large number of interconnections and variability over time. Social relationships

✉ Antonio Zinilli
antonio.zinilli@ircres.cnr.it

1 Sapienza University of Rome, Rome, Italy

2 IRCRES CNR, Rome, Italy

among agents are an example of complexity, where nodes can be people, programs, projects etc., and links represent their interactions (Watts et al. 2002; Arenas et al. 2004). This paper studies the collaboration among researchers in a specific Italian funding program, the Projects of National Interest (PRIN), which aims to support the academic research (Reale and Zinilli 2014). Through the PRIN, which represents an important governmental competitive tool to give resources for collaborative research in different interest sectors, we want to analyze the factors behind the collaborations among researchers, during the period from 2000 to 2011. The main aims of this paper are to understand the distribution of the connections among researchers who decide to collaborate on project funding application and why the links in the research collaborations, given a set of covariates, change over time. This paper does not considered the general collaboration among scholars in Italy, but only those collaborations that have received funding in the PRIN context. Since researchers join or leave the network and consequently the network composition changes over time, it is important to understand the networks formation according to a set of attributes, which have been selected on the relevant literature. We have studied a competitive funding scheme through dynamic network analysis techniques, as there is more and more a tendency towards competitive systems regarding project funding (Lepori et al. 2007). We think a power-law distribution exist with the network collaborations, there is a spatial dimension that influences the ties, there is no tendency to collaborate with researchers from other disciplines and we think that there is a dissimilarity in the H Index in the collaborations. The paper is organized mainly in five steps. First, we proceed with an analysis of the literature on network collaborations. Secondly, we check if there is the presence of a power law in our dataset in the four areas of interest (chemistry, physics, economics and sociology). Thirdly, a model is developed to analyze how collaborations are formed in the four areas over time. A stochastic actor-oriented model is used in the analysis to estimate parameters which are at the base of the mechanism of change. In this model, we take into account different PRIN calls from 2000 to 2011, putting into the model a set of control variables. Then, we present the results of estimates of the longitudinal analysis. Finally, we check our hypothesis and present the conclusions that derive from the estimates.

## Conceptual framework

The current literature focuses on the effectiveness of the approach to examine network models and patterns of collaboration in scientific communities and to describe the different roles of researchers in the network (e.g. more isolated or central researchers or those who have a brokerage role). In the last years, research collaborations have been studied through several approaches: one approach tends to investigate how we can measure research collaboration (a methodological approach); another approach tends to study the impact of collaboration on productivity (among researchers, between firms and university etc.); another approach studies what are the indicators that drive the formation of research collaborations, taking into consideration patterns of geographic proximity and scientific interaction levels (Andersson and Persson 1993; Newman 2001a, b; Katz and Martin 1997). Scientific collaboration is a set of informal functions (e.g. face-to-face contacts) and formal activities (e.g. participating in research projects or professional conferences) among scientists involved in producing knowledge (De Stefano et al. 2013). Usually, researchers start to collaborate with others through personal contacts developed during their career,

often begun at professional conferences (Laudel 2006). Most of the studies focus on co-authorship networks or co-citation/citation networks. According to Gmür (2003, p. 27–28) the number of co-citations represent an indicator of proximity in terms of content. Hummon and Doreian (1989) have developed new methods (main-path algorithms) to weigh the links in a citation network of articles. Hummon and Carley (1993) used bibliographic citations and the SNA technique of path analysis to study the connections between researchers published in the journal Social Networks. In this paper we focus on collaborations that were born in projects funded on a competitive basis in four disciplinary fields.

An important law (especially in the literature of physical sciences) to model the overall mechanism of the formation of the network is the power-law degree distribution, which provides a mechanism for network growth tied to the fact that the new vertices tend to connect the vertices that already possess a high degree of connections. In particular, when the probability of being connected to a given vertex is proportional to its degree, such distribution results can be observed (Albert and Barabási 2002). Of course, this mechanism may occur with different variations in different real contexts. The power-law is a mathematical distribution of the concept of the Matthew effect described by Merton (1968), which is a possible effect of another concept used in network science, that is a preferential attachment (Newman 2001a, b). In the scientific collaboration networks, the power law distribution has been observed more times (Newman 2001a, b), which means the more connected nodes are destined to have much more connections in the future than the new nodes in a network. Many empirical studies have also shown distributions of power law at the upper tail of a distribution (Sinha 2006), other studies have shown that income and wealth distributions follow log-normal distributions (Chatterjee et al. 2005). However, as explained by Clauset et al. (2009) detecting a power law distribution in empirical data may be very difficult and for this reason it is important to compare the power law model with some other distributions (Newman 2005).

Very often junior researchers are forced to join senior researchers to join a network. Following this, the first question is:

**Q1** is there a power law distribution in the collaborations in a competitive funding context?

**H1** the hypothesis is that exist a power-law distribution to the network collaborations in competitive systems.

The everyday use of the word collaboration suggests researchers working together to achieve a common goal (most likely conducting research and developing new scientific knowledge) (Katz and Martin 1997; Newman 2001a, b, 2004). Funding programs are always keen "to increase the level of collaboration engaged in by the researchers whom they support in the belief that this will bring about better benefits", increasing researchers' chances of funding success (Katz and Martin 1997, p. 2; Laudel 2006; Beaver 2001). We are aware that the collaboration pattern is the result of a complex interplay of other factors. Boschma (2005) suggests five forms of proximity: Geographical proximity, cognitive proximity, organizational proximity, social proximity, institutional proximity. Therefore, the decision to collaborate with other researchers is influenced by different dimensions of proximity, which may overlap, reducing the independence concept. Usually, the several concepts of proximity are applied to inter-firm collaborations, innovation and regional economic development (Anand and Khanna 2000; Boschma 2005). Fernández and López (2015) examine the effect of geographical, cognitive, institutional, organizational and social proximity on scientific collaboration among academic institutions using both

bibliometric data on co-authorship and EUMIDA data on institutional-level information. The results show evidence of the importance of geographical, cognitive, institutional and social proximity in the rise of research collaborations while the effect of organizational proximity seems to be weaker.

The spatial proximity (or geographical) is an important determinant of collaboration patterns. Indeed many collaborations are begun in informal contexts, which geographical proximity facilitates (face-to-face meetings are easier to organize) (Andersson and Persson 1993; Katz 1994; Boschma 2005). Cognitive proximity refers to the degree of similitude of the knowledge bases of organizations and institutions (Nooteboom et al. 2007). Organizational proximity is defined as the degree of strategic interdependence between two organizations, and it reduces the uncertainty about the behaviour of the future partner (Boschma 2005). Social proximity is defined by the existence of direct and informal personal interaction of the employees or the managers of two different organizations (Boschma 2005; Boschma and Frenken 2009). Institutional proximity is defined by the similarity of informal and formal rules shared by actors (North 1990).

In this work we explore the possibility of using geographical proximity and cognitive proximity to understand how they influence the choice of partners in the PRIN context.

In particular, researchers nearby are more likely to collaborate in the PRIN. Researchers whose universities are far away are less likely to collaborate together, even if their research interests are close. The literature on the contribution of spatial distance has analyzed the importance of distance in time (Scherngell and Lata 2011; Smith and Katz 2000). Therefore, we look at the geographical dimension of researchers that have collaborated in the PRIN program by examining the correlation between proximity and the decisions to collaborate with different partners. The question that we ask is:

**Q2** is there spatial correlation in the formation of collaborations in a competitive funding context?

**H2** the hypothesis is that spatial dimension positively influences the evolution of the collaboration network.

In our case we can mean the cognitive proximity in terms of disciplinary, measured through the scientific area of belonging of the researchers. The disciplinary proximity promotes communication and the exchange of knowledge among researchers of the same scientific area.

Gibbons et al. (1994) note that modern science is adopting new organizational forms largely using interdisciplinary perspectives instead of disciplinary ones. Different works have studied collaboration when there is heterogeneity of knowledge and skills (Durfee et al. 1989; Andersson 2011). Increasingly, there are calls for more interdisciplinary approaches to research, along with encouragement for greater collaboration among researchers (Hicks and Katz 1996; Gibbons et al. 1994). The paper investigates the disciplinary proximity in terms of the researchers affiliated area, if two researchers belong to the same area they are close if not the opposite applies (similar knowledge bases).

The research question is:

**Q3** is there a tendency to collaborate between individuals from different disciplinary backgrounds in the changing links?

**H3** the hypothesis is that in the PRIN there is not a tendency to collaborate with scholars with different knowledge.

For chemistry and physics we have the h-index[1] for each researcher. In this paper we use this measure to investigate how the variation in individual scientific impact is related to the collaboration network. We consider the h-index as prestige, with all its limitations (Bornmann and Marx 2011). The citation could be viewed like a vote, with the same meaning has an academic context (Davis 2008); the status of a researcher in a social context is given by the total number of citations by other researchers (Bollen et al. 2006). The question is:

**Q4**   is there an h-index assortativity[2] in the choice of partners?

**H4**   the hypothesis is that there is not an assortativity dimension, but a preferential attachment about scientific prestige.

Assortativity dimension reflects tendencies for researchers with high h-index to preferably be tied to other researchers with a high h-index. We think that is exactly the opposite, those researchers with a low h-index tend to join with those who have a high h-index (preferential attachment behaviour). We calculated the h-index using all years of study, this information was taken from Scopus search.

## Dataset and methodology

The paper uses two approaches to study the dynamic complex networks: a first interest is to identify the observed distribution of links among researchers in the four areas of interest through distribution models, then we will use a stochastic model to understand how the links change over time. In order to do this, we have built through the PRIN website a large and novel dataset containing 4322 researchers from 98 universities and research institutes that have been selected for PRIN allocation from 2000 to 2011. We chose this time period because it was long enough for a longitudinal analysis and because in subsequent years the data was not available.

The sample refers to four research areas: chemistry, physics, economics and sociology. Both physics and chemistry are characterized by consolidating collaborative patterns and publication habits in international indexed journals. As to the social sciences, the focus is on areas where collaborations are not so diffused between scholars and the use of indexed journals is not widespread, since scholars prefer diverse types of output production (e.g. books). The selection of the aforementioned areas comes from the need to have a sample representing different characteristics of the epistemic communities. This study concerns an Italian funding scheme and it cannot be generalized to the entire academic community, therefore we can consider this work as a proxy of the overall communities.

12 years of PRIN program are observed. Originally, the dataset was in two-mode network (researchers by project), then we built the dataset in one-mode through a command implemented in STATA (Zinilli and Cerulli 2015). We consider the PRIN research team membership as the most viable way of collecting data on collaboration for our analysis. Data collected on PRIN grants allocated in the considered period allow us to understand the characteristics of the proponents. In fact each research team has a principal

---

[1] The h-index is a measure used to indicated the impact of a researcher based on how often his/her publications have been cited.

[2] With assortativity effect we mean the tendencies for researchers with high h-index to preferably be tied to other researchers with high h-index.

**Table 1** Variables used in the SAOM model

| Name of variable | Description of variable | Type of variable |
|---|---|---|
| H index | Only for chemistry and physics, we calculated the h-index based on all the papers attributable to each author, we took these information from Scopus search | Discrete data |
| Size university | Small universities are those that have up to 500 researchers, middle-sized are those that have between 500 and 1500 researchers and large universities are those that have more than 1500 researchers; the size is expressed by disciplinary area (e.g. physics, chemistry). Division by the size of the university was taken from Evaluation of Research Quality (VQR) for the 2004–2010 (source: http://www.anvur.org/rapporto/) | Small/medium/large |
| Academic role | This variable indicates the role by researcher for each year within the affiliation institute. This information was taken from the PRIN site | Researcher/Associate professor/Full professor |
| H index × role | This variable has been created with the aim of isolating the presence of a modification of the effect of H index determined by seniority | H index x Academic role |
| Gender | This is a control variable | Male(1)/Female(0) |
| Geographical area | This variable is based on ISTAT NUTS classification. Groups of regions observed at five different levels, obtained by ISTAT (Italian National Institute of Statistics): North-West, North-East, Centre, South and Islands. If a researcher comes from North-West and the partner from Centre we assigned value 1, because they came from two different geographical area | ISTAT NUTS classification (0 = same geographical area/1 = other area) |
| Interdisciplinary | We built this variable on the base of the scientific area-field of specialization. If two researchers belong to the same area we gave value 0, in the other case | Dummy (0 = same scientific area/1 = others) |
| Netchange (Call PRIN) | We built this variable on the base of the "Calls" because since 2005 it has not been possible to participate to the PRIN for two consecutive years (value 0 before 2005 and value 1 after 2005). We considered this effect with control variable on the changing of connections.in the SAOM model | Dummy (0/1) |
| Collaboration (Call PRIN) | We built this variable on the base of the "Calls" because until 2004 the guidelines by government about the number of components for each group were the same, after this date the guidelines changed. We considered this effect with control variable on the changing of connections in the SAOM model | Dummy (0/1) |
| Funding (Call PRIN) | We built this variable on the base of the "Calls" because since 2009 applications for funding have changed. We considered this effect with control variable on the changing of connections in the SAOM model | Dummy (0/1) |
| Gender (Call PRIN) | We built this variable on the base of the "Calls" because since 2007 the Call has required gender equality. We considered this effect with control variable on the changing of connections in the SAOM model | Dummy (0/1) |

investigator and one or more responsible for the research units, the gender, academic position, scientific quality, role in the PRIN, affiliation (we have taken into account that researchers change institutions during the 12 years under analysis).

To study how the collaborations are formed over time, the package RSiena for "R" environment is used in the analysis. In four models that we have built (a model for each area) we have included a number of control variables (some labelled like "Call PRIN") to understand how government guidelines have influenced changes over time. We put these variables because the exclusion of control variables would compromise the internal validity of the model and to understand if there is a policy's effect in the changing link. Initiatives and directives by government could lead to the establishment of contacts with people other than the scientifically most interesting ones. A variable of interaction between h index and the academic role has been added, because there could be the possibility of a confounding role by researcher's seniority (is there a modification of the effect of H index due to seniority?). The goal is to isolate the presence of a modification of the effect of H index determined by seniority.

Variables used in the model are (Table 1):

The PRIN dataset contains a changing composition, indeed the network can acquire new researchers over time (joiners), or lose the researchers (leavers). As Bellotti (2012) explains the PRIN calls have changed since 2005 and after this date it has not been possible to participate to the PRIN for two consecutive years. To remove problems of convergence of the estimation algorithm, we had to combine the date for years 2004 and 2005, 2006 and 2007, and 2008 and 2009. Furthermore, we have applied composition change directives in the RSiena package. Research joiners and leavers during the years could have influenced the model, although parameters are estimated with careful consideration through imputation techniques recommended by Huisman and Snijders (2003) (Huisman and Steglich 2008). Before moving forward in the analysis we introduce some basic network terminology. A network is composed of nodes (vertices), in our analysis the researchers, and the links (edges), belonging to the same research team. It is represented as a $n \times n$ adjacency matrix, where the element $a_{ij}$ provides information of the existence of a link from researcher $i$ to researcher $j$; the network is undirected.

## Power-law distribution

In previous years different papers on real complex networks analyzed a self-organizing processes in the real networks.

The network is undirected, an actor's degree is the number of other actors to which it is directly connected. For degree we mean the number of co-workers with which each researcher collaborated along the years (repeated collaborations with the same partner are counted only once).

Degree distribution is the probability that a randomly chosen node has $x$ connections. The probability $P(x)$ that a node in the network interacts with $x$ other nodes decays with the law:

$$\mathrm{P}(x) \propto ax^{-\gamma}$$

Where $P(x)$ is the probability to encounter value $x$ and $\gamma$ as the scaling exponent ($\gamma \succ 0$). Scale-free networks are open and dynamically formed by the continuous addition of new nodes. The likelihood to have a node with $x$ links is function of the number of links 'scaled' for the exponent on which the skewness of the distribution depends. It can happen that

power-law distribution can look like other types of distributions (for instance to exponential distribution) or the right-skewed distributions can be better fitted with a power law with exponential cut-off (Giot et al. 2003). In other cases the distribution follows a power-law only over a lower bound ($x_{\min}$) that has to be estimated when fitting the empirical data. Power-law has two important features: (1) it does not have a peak at its average value and it starts at its maximum value and then decreases all the way to infinity; (2) the rate at which the power-law decay is much slower than the decay rate of other distributions (for example with respect to an exponential distribution) and this brings to a much greater likelihood of extreme events. In order to estimate $x_{\min}$ we have followed different procedures (Clauset et al. 2009). Below (Table 2) the distributions we have used to compare with power-law, for each distribution we indicated the basic functional form f(x):

First, to understand if we can speak of a power-law we follow a straight line on logarithmic axes (log–log plot). Through an observation of the log–log plot of the probability distribution function we can see if there are values that change and don't follow the power law distribution. The slope is the value of the scaling parameter $\gamma$ and a slope changing reveals the lower bound when there is not a power-law for all its values.

A power law is a linear relationship between logarithms, of the form:

$$\log P(x) = -\gamma \log x + \log a$$

In the continuous case, the constant $a$ in the equation is provided by the request of normalization, namely:

$$\int_{x_{\min}}^{\infty} p(x)dx = 1 \rightarrow a = (\gamma - 1)x_{\min}^{\gamma-1}$$

In discrete case, as in this study, we have:

$$p(x) = 1 \rightarrow \alpha = \frac{1}{\zeta(\gamma, x_{\min})}$$

with:

$$\zeta(\gamma, x_{\min}) = \sum_{n=0}^{\infty} (n + x_{\min})^{-\gamma}$$

it is clear that $\gamma > 1$ otherwise the equation diverges. To have more consistent estimates, using the methodology by Clauset et al. (2009) to measure power-law behaviour, we have compared it with some other distributions. To fit a discrete power law to this data and to understand if there are alternative models fit the data better, we have calculated the $x_{\min}$ (through the Kolmogorov–Smirnov statistics) that minimizes the distance between empirical data and the power-law model which best fits it. For a given value $x_{\min}$, the scaling parameter is estimated by numerically optimising the log-likelihood (Gillespie 2015). For continuous cases, the optimization takes place through a maximum likelihood estimator:

**Table 2** Functional form by distribution

| Distribution | f(x) |
| --- | --- |
| Log-normal | $\frac{1}{\sqrt{2\pi}\sigma x}e^{-(\ln x - \ln x_0)^2/(2\sigma^2)}$ |
| Exponential | $e^{-\lambda x}$ |
| Poisson | $\mu^x/x!$ |
| Power-law + cut off | $x^{-\alpha}e^{-\lambda x}$ |

$$\hat{\gamma} \simeq 1 + n \left[ \sum_{i=1}^{n} \log\left( \frac{x_i}{x_{\min}} \right) \right]^{-1}$$

where $x_i = 1,2,3\ldots$, n are observations such that $x_i > x_{\min}$. In discrete cases, we can estimate $\hat{\gamma}$ in different ways, one way is through the numerical maximization of the logarithm of the likelihood function (Clauset et al. 2009):

$$L(\gamma) = -n \log \zeta(\gamma, x_{\min}) - \gamma \sum_{i=1}^{n} \log x_i$$

The standard errors for estimated parameters are computed with standard bootstrap methods with 5.000 replications. Then, a goodness of fit tests is applied and used to compare power-law fits with fits of other alternatives. A goodness of fit tests shows us how good is the model with the parameters $x_{\min}$ and $\gamma$ estimated before. As Clauset et al. (2009) have shown the hypothesis is tested using a bootstrapping procedure. If the $p$ value generated is large, any difference between the empirical distribution and the model can be explained from statistical fluctuations. If $p \simeq 0$, the model does not offer a reasonable fit to the data and other distributions could be more appropriate. Following the literature the power-law hypothesis is rejected with a $p$ value <0.1. On the other hand, a large p-value does not mean that power-law is the best model and other alternative hypotheses could be better, such as the log-normal, the exponential or Poisson. To compare these distributions, the logarithm of ratio of the likelihoods of data for two different models are calculated, if the value is positive the power law model is preferable, if negative it is preferable to the other distribution. To understand if the sign of the value is significant, a standard technique is to use the Vuong's test (Vuong 1989). Lower $p$ values tell us that the sign is a good indicator of which model better measures empirical data. Instead, higher $p$ value indicates that we are not able to decide which model is preferable.

## Stochastic actor-oriented model

Stochastic actor-oriented models (SAOM) have been built to study the complexity of network panel data and thus to model change in social networks. The first study with SAOM of social networks panel data was introduced by Snijders (1996). In SAOM, the dependent variables are actors' choices about changes of network relations. Each observation is represented by an $n \times n$ matrix $x = (x_{ij})$ where $x_{ij}$ represents the link from the researcher $i$ to the researcher $j(i, j = 1, 2, \ldots n)$. The independent variables can be endogenous or exogenous. Endogenous variables explain network changes with existing network structures: an example of endogenous structure is the transitivity or reciprocity. These types of variables capture the influence of specific local network structures of co-collaboration on the probability of creating a new tie. Exogenous variables are the actor covariates, which capture the influence of individual characteristics of researchers on the probability of creating a new tie, for example same sex, similarity in a specific characteristic, etc. In the SAOM each change in the network is made according to processes of individual choice. This important assumption is realistic for what we are studying.[3] Into the

---

[3] In the models we include variables at a structural level, as well as also control variables.

stochastic actor-oriented model are embedded Markov processes,[4] random utility function and Monte Carlo simulation. The model assumption is that the discrete network observations $(t_1, t_2,…t_n)$ are only snapshots of an unobserved underlying dynamic sequence, the evolution between $t_{n-1}$ and $t_n$ is assumed to be continuous and simulated with a Monte Carlo method. These models for network evolution are outcomes of a Markov process evolving in continuous time, which a special stochastic process. A continuous-time Markov process determines the change of the network connections. Given random variables $X_t$, with $t \in (0, \infty)$ in a discrete set $S$, to extend the notion of Markov chain to that of a continuous time Markov chain we require:

$$P\big[X_{s+t} = j|X_s = i, X_{sn} = i_{n,…,}X_{s1} = i_1\big] = P\big[X_{s+t} = j|X_s = i\big]$$

for all $t > 0$, $s > s_n > … > s_1 \geq 0$ and $i, j, i_k \in S$. The quantities $P[X_{s+t} = j|X_s = i]$ is a transition probabilities. This is the analogue of the Markov process with discrete time variable, just that here there is continuous parameter (in this model simulated with a Monte Carlo method).

The idea behind this type of model is modelling the change process through two important components: the change opportunity process (rate function) and the change determination process (objective function).[5] The actors are in control of their links and they each seek to change these links (i.e. create, maintain or dissolve the links) such that their personal "satisfaction" with the network composition is maximized. An important assumption in this sense is that the actor has full knowledge of the network, including the other actors, links and their characteristics. The opportunities to change a link are modelled according to a Poisson process with rate $\lambda_i$ for each actor $i$. The interpretation is that actor $i$ to change one of the tie variables occur at a rate of $\lambda_i(x_0)$, where $x_0$ identifies the state of the network at a certain time. In the basic model, all the actors have the same opportunity for change, which is equal to a constant parameter $\lambda_i = p_m$. When we include individual attribute $(v_i)$ or structural variables as the degree $(\sum jx_{ij})$ we introduce the heterogeneity in change opportunities. In these more complex models the rate function is given by:

$$\lambda_i\big(x^0, v\big) = p_m \exp\Big(\alpha_1 v_i + \alpha_2 \sum_j x_{ij}\Big)$$

From the general theory of continuous-time Markov chains (Norris 1997) - following on a current state $x^0$, and a set of permitted new states $C(x^0)$ and the product of the two model components $\lambda_i$ and $p_i$ (with $p_i$ defines the probability distribution of choices)—follows the existence of the intensity matrix that describes the rate at which $X(t) = x^0$ tends to transition into $\tilde{X}(t + dt) = \tilde{x}$ as $dt \rightarrow 0$:

$$q_{(x^0,x)} = \lim_{dt \to 0} \frac{P\Big\{X(t + dt) = x \mid X(t) = x^0\Big\}}{dt} \big(x^0 \neq x\big)$$

where $x \in X$, $q_{(x^0,x)}$ whenever $x_{ij} \neq x_{ij}^0$ for more than one element $(i, j)$ and $q_{(x^0,x)} = \lambda_i(x^0, v, w)p_i(x^0, x, v, w)$ for graphs $x$ and $x^0$ which differ the element with index $(i, j)$. Given that an actor $i$ has the opportunity to change a relation, the choice for this actor is to change one of the link variables $x_{(i,j)}$, because actors can only change one link variable at a

---

time. Changing the link variables $x_{(i,j)}$ will lead to a new state $x$, with $x \in C(x^0)$. In order to model choice probabilities, a classical multinomial logistic regression specified by an objective function $f_i$ is used (Snijders et al. 2010):

$$p\{X(t) \text{ changes to } x | i \text{ has a change opportunity at time } t, X(t) = x^0\}$$

$$= p_i(x^0, x, v, w) = \frac{\exp(f_i(x^0, x, v, w))}{\sum_{x' \in C(x^0)} \exp(f_i(x^0, x', v, w))}$$

When actors have the opportunity to change their relations, they choose to form a new tie to try to maximize their objective function $f_i$. This objective function describes preferences and limits of nodes (our case the researchers). Precisely, relationship choices are then determined by a linear combination of effects, depending on the current state $(x^0)$, the potential new state$(x)$, individual attributes$(v)$ and attributes at a dyadic level$(w)$. Effects related to the current state of the network are endogenous, implying a self-reproduction of network structures, such as transitive closure or betweenness effect. Individual attributes are effects modelling the propensity of some nodes to create more links. Dyadic effects indicate the propension of actors with analogous characteristics to form links.

Parameters indicating the strength of each effect in determining the network dynamics are estimated from longitudinal data using a simulation-based approach inspired by the method of moments. The solution of the moment equation is obtained by a variation of the Robbins and Monro (1951) algorithm (for further information see Snijders 2001). The stochastic algorithm has the task to simulate the development of the network and estimates the parameters that minimize the distance between observed and simulated networks. During the iteration phase, the provisional parameters of the probability model are gradually adjusted in a way that the simulated networks fit the observed networks. Then the parameters are held constant to its final value, with the aim to evaluate the goodness of fit of the model and the standards errors.

Finally, the parameters estimate by SAOM can be read as non-standardized coefficients obtained from a classical logistic analysis. Indeed, the parameter estimates are log-odds ratio, and they can be interpreted like the log-odds of ties change with one unit change in the relative independent variable.

## Empirical results

In this section we show the results of the different research questions. As a first step we want to visualize our data in order to get an idea of how the data looks. The table below presents some descriptive statistics of the network structure, where the values are normalized (Table 3).[6]

In the PRIN, the nodes (researchers) and the number of connections in the natural sciences are greater than sociology and economics, consistent with the different size of the community. The average distance is a measure of how far a researcher is from the other researchers; in physics, there is a smaller distance than the other areas. Indeed, the average geodesic distance between researchers in the network is 5.94. Clustering coefficient is high for all disciplinary fields, this coefficient goes from 0 to 1. In PRIN projects each researcher is linked to everyone else within the project and for this reason the clustering

---

[6] This because over time the government' guidelines about collaborations are changed.

**Table 3** Descriptive analysis

| Area | Nodes (2000–2011) | Number of links (2000–2011) | Maximum geodesic distance (Diameter) (2000–2011) | Average geodesic distance (2000–2011) | Average clustering coefficient (2000–2011) |
|---|---|---|---|---|---|
| Chemistry | 1314 | 5923 | 28 | 9.84 | 0.83 |
| Physics | 1185 | 4426 | 18 | 5.94 | 0.74 |
| Economics | 1107 | 3546 | 24 | 9.84 | 0.79 |
| Sociology | 716 | 2418 | 20 | 8.03 | 0.78 |

*Source* Author elaboration

coefficient, average geodesic distance and maximum geodesic distance have been calculated on a whole network and not year by year.[7] The mean clustering coefficient is measured by the fraction of paths of length two in the network that are closed. The "small world" phenomena, the combination of short average path lengths together with a high cluster coefficient, seem to be present in physics. This phenomenon happens because a researcher acts as a shortcut between distant researchers. A possible explanation is because in physics more researchers have had collaborations of research through past collaborations with common third researcher. We expect that the closest nodes are those that work together in large laboratories (they belong to the physics' experimental sector). On the other side the most isolated nodes seem to be those belonging to the theoretical physics, that have less need to collaborate (verified observing the nodes more in depth[8]).

After this brief description of the states of the network on the four disciplinary fields, the power-law distribution is presented. The first hypothesis argues that researchers' network follows a power-law distribution. Given the definitions above (paragraph "Power-law distribution") a power law is a linear relationship between logarithms. This plot should follow a straight line in case of a power law. Figure 1 presents the log–log plot of chemistry:

The log–log plot shows a straight line only after a specific value, suggesting that a power-law model will have a good fit after a minimum value. Chemistry follows a power law from a $x_{min}$ equal at 5 and with an exponent equal at 2.7. For physics $x_{min}$ is equal at 9 with a exponent of 3.83 and sociology with a $x_{min}$ equal at 12 with an exponent of 2.11. In economics the minimal Kolmogorov–Smirnov value is reached keeping the lower bound at 12 with a scaling exponent of 2.04. The power-law p-value (second column in Table 4) is not statistically significant for physics, economics and sociology. For that reason alternative distributions should be explored.

Chemistry seems present a pure power law with a *p* value of 0.62 (thus > 0.1): physics, economics and sociology don't have a power law but follow other types of distributions. From Table 4 we can see that the physics, economics and sociology are not fitted well by the power-law model according to the goodness-of-fit test used. Physics follows a log-normal distribution, sociology data fit best to an exponential distribution and economics there is not a statistical significance. In conclusion, we find that in the PRIN program only

---

[7] If we had calculated the cluster coefficient per year it would have the value 1 because within each project everybody is linked to everyone else.

[8] Centrality indices were calculated (degree, betweenness and closeness centrality) and we noticed that scientists with lower centrality indices were theoretical physicists.
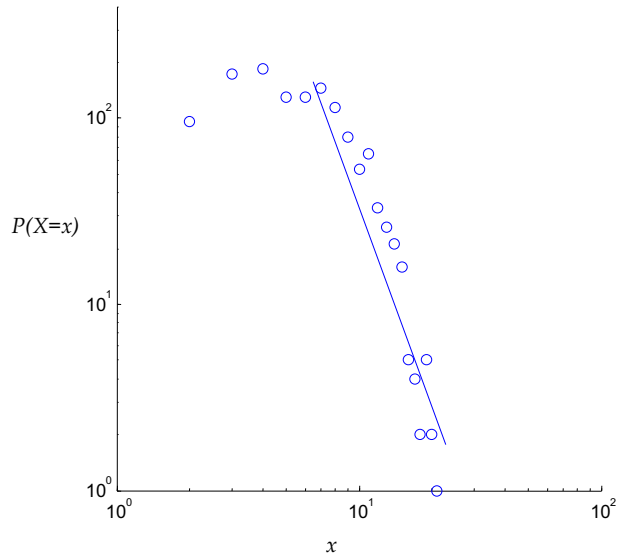
**Fig. 1** Log–log plot of chemistry



**Table 4** Power-law and alternative distributions

| Degree | Power law | Log-normal | | Exponential | | Poisson | | Power-law + cut off[a] | |
|---|---|---|---|---|---|---|---|---|---|
| | p* | LR | p* | LR | p* | LR | p* | LR | p* |
| Degree chemistry | 0.62 | −0.83 | 0.20 | −1.22 | 0.11 | −0.37 | 0.36 | −1.12 | 1.17 |
| Degree physics | 0.01 | −3.57 | 0.01 | −5.03 | 2.43 | 0.97 | 0.84 | −0.35 | 0.05 |
| Degree economics | 0.03 | −0.34 | 0.36 | 2.29 | 0.02 | 2.57 | 0.01 | −2.25 | 0.22 |
| Degree sociology | 0,06 | −1.10 | 0.15 | −1.71 | 0.04 | −0.79 | 0.21 | −1.32 | 0.96 |

[a] The pure power-law model is nested within the power-law with exponential cut-off model, and for this reason the latter always provides a fit at least as good as the former; therefore the LR for power-law with exponential cut off will be always negative or zero

*Source* Author elaboration

the network collaboration of chemistry is fit well by a power-law model between 2000 and 2011, even if $p$ values of other distributions (in chemistry) are so large that the tests cannot exclude better distributions.

## SAOM results

Following the approach by Snijders (2005) initially we included in the model only basilar network effects, in order to check for endogenous dynamics.[9] At a later time, we add further effects and dropped those that were not significant. In this way it was possible to avoid model instability when running the algorithm and getting reliable estimates of the

---

[9] The model was built selecting only some effects which reasonably drive the evolution of the network among the several effects defined and provided in the package RSiena, through the command "ef-fectsDocumentation". We chose the ones best suited to our studio and network characteristics.

parameters. Consequently, the first model presented (Model 1) is very simple and only includes density, transitive triads, betweenness and the degree assortativity effects. The density must be included in all objective functions. It is a sort of intercept of the model and represents the tendency to form arbitrary edges. Transitivity indicates that two researchers sharing a tie towards a third researcher are more likely to engage in collaborative activities between them than to other researchers. This effect measures the tendency of researchers to cluster together. Betweenness effect measures the intermediation dynamics in the evolution of the network which are present if researchers tend to stay in the middle between indirectly connected couples of researchers. The degree assortativity effect reflects tendencies for researchers with high degrees to preferably be tied to other actors with high degrees.

For network evolution, the rate function describes the average number of changes in network ties between measurement points. Below we show the estimates for the chemistry area (Table 5):

The table above shows that the average rate of change in the final model from 2000 to 2011 is 3; it means that between 2000 and 2011 on average the expected number of changes was 3 for each researcher. The density effect is negative and significant; this variable indicates that there is an opportunity cost in the establishment of each relation. Thus, the tendency of researchers to start a relationship is driven by other variables, compensating for this cost.

The significant transitivity effect indicates a tendency for transitive closure (specifically, the transitivity effect indicates a preference for collaborating with friends of friends); we could say there is trust in sharing common partners of research.

Betweenness effect in the second model is not significant excluding the tendency to intermediation by researchers.

Degree assortativity does not seem to affect the actors' choice about changes of network relations.

Spatial proximity seems to be a relevant variable, it means that there is a tendency to create collaborative ties based on the location in the same geographical area.

The h-index effect, a positive parameter implies that researchers prefer ties to other researchers with different values on h-index variable (preferential attachment about the prestige). The model says that the effect of the interaction is not statistically significant, therefore the seniority does not change the effect of the H index.

The changing of the funding specification and the thresholds in the number of collaborations in the PRIN call seems to influence the partners' choice.

Below the results of physics (Table 6).

In physics the average rate of change is slightly lower than in chemistry; therefore, physicists have changed less partners in these years. The variables of control "Funding" and "Collab" seems to have an effect on the choice of partners. The transitivity effect, h-index and geographical proximity are significant like in chemistry. Even in this case the variable of interaction between H Index and Academic role is not statistically significant, therefore the seniority does not change the effect of the H index on changing of ties. The important difference compared to chemistry is the role played from the endogenous variable "betweenness". The Betweenness effect in the second model is significant, in this situation the only node that is connected to both nodes as an advantaged for it can directly acquire resources from them and manage the flows among them. In physics there are nodes that can play as intermediaries among indirectly connected couples of nodes.

The last two tables show the SAOM result for economics and for sociology (Tables 7, 8).

**Table 5** SAOM for chemistry

| Parameters | Model 1 | | Model 2 (Final) | |
|---|---|---|---|---|
| | Estimate β | t-ratios | Estimate β | t-ratios |
| Rate (2000–2011) | 3.150 (average all period)[a] | | 3.009 (average all period) | |
| Density | −2.610*** (0.104) | −0.033 | −1.944** (0.037) | −0.023 |
| Transitive triplets | 0.210*** (0.356) | 0.025 | 0.440** (0.022) | 0.044 |
| Betweenness | −0.817* (0.616) | −0.048 | −0.002 (0.729) | −0.078 |
| Degree assortativity | −0.728 (0.771) | −0.055 | | |
| Size university | | | −0.997 (0.907) | −0.390 |
| Collab (Call PRIN) | | | 0.008 (0.044)** | 0.090 |
| Funding (Call PRIN) | | | 0.643** (0.064) | 0.030 |
| Gender (Call PRIN) | | | −0.010 (0.089) | −0.701 |
| Interdisciplinary | | | −0.691 (1.680) | −0.136 |
| Netchange (Call PRIN) | | | −2.021 (1.064) | −0.090 |
| Same academic role | | | 0.002 (0.207) | 0.057 |
| Gender | | | 0.128 (1.063) | 0.101 |
| H Index | | | 0.175** (0.014) | 0.038 |
| H Index × role | | | 1.277 (1.022) | 0.558 |
| Geographical proximity | | | −0.115** (0.012) | −0.011 |

"Estimation" = average of parameters' estimations. Standard deviation in parenthesis; "t-ratios" = test for the convergence of the algorithm ($t$ values <0.1)

[a] For reasons of space we report the average of the parameters (from 2000 to 2011)

Significance of the estimation values (probability of acceptance of the null hypothesis): *** <0.01; ** <0.05; * <0.10

In economics the average rate in the final model in considered years is 3.7; it means that on average the expected number of changes is almost 4 for each researcher. Even here there is an influence of the transitive effect and geographical proximity on the changing of links. In addition to the dummy variable "Funding", in economics the effect of control variable is added, here called "collab[10]", which indicates a change in the guideline by MIUR about the possibility of participating in the PRIN program in two consecutive years. In this context we can only say that there is an influence of this variable in the change of the links, but we cannot put this variable in relation with the other variables of the model.

Below the results for sociology.

In sociology the average rate of change is equal to 4.2, it means they tend to change more than other disciplines. Also here there is a transitive effect and a geographical effect like in economics. What differentiates sociology are the variables "interdisciplinary" and "same academic role variables".

The negative sign of variable "interdisciplinary" means that there is a low tendency to collaborate with those researchers who come from other disciplinary fields. The positive sign of variable "same academic role" means that in sociology, in the PRIN context, a researcher tends to change the links on the base of academic level; for instance, a

---

[10] The PRIN calls have changed since 2005 and after this date has not been possible to participate to the PRIN for two consecutive years.

**Table 6** SAOM for physics

| Parameters | Model 1 | | Model 2 (Final) | |
|---|---|---|---|---|
| | Estimate β | t-ratios | Estimate β | t-ratios |
| Rate (2000–2011) | 3.322 (average all period) | | 2.856 (average all period) | |
| Density | −1.377** (0.033) | −0.445 | −1.124** (0.021) | −0.036 |
| Transitive triplets | −0.025** (0.056) | −0.028 | 0.089** (0.002) | 0.085 |
| Betweenness | 0.127** (0.059) | 0.012 | 0.122** (0.095) | 0.019 |
| Degree assortativity | −0.364 (1.922) | −0.154 | | |
| Size university | | | −0.352 (0.907) | −0.012 |
| Collab (Call PRIN) | | | 0.438 (0.564)* | 0.041 |
| Funding (Call PRIN) | | | 0.201*** (0.002) | 0.047 |
| Gender (Call PRIN) | | | −2.411 (0.554) | −3.051 |
| Interdisciplinary | | | −2.322 (1.731) | −5.032 |
| Netchange (Call PRIN) | | | −1.481 (0.064) | −0.450 |
| Same academic role | | | 0.121 (0.002) | 0.123 |
| Gender | | | 0.125 (1.801) | 3.332 |
| H Index | | | 0.075** (0.114) | 0.033 |
| H index × role | | | 0.845 (0.173) | 0.192 |
| Geographical proximity | | | −0.010*** (0.312) | −0.021 |

"Estimation" = average of parameters' estimations. standard deviation in parenthesis; "t-ratios" = test for the convergence of the algorithm ($t$ values <0.1)

Significance of the estimation values (probability of acceptance of the null hypothesis): *** <0.01; ** <0.05; * <0.10

researcher prefers to do research with a full professor instead of another researcher (preferential attachment on the base of academic role). This is a feature found only in sociology.

## Discussion and conclusions

We analyzed data on PRIN funded grants to figure out the collaboration patterns among researchers in a competitive context. We have noticed that there are different types of distributions in the four networks highlighting the distinct scientific traditions by the four academic communities. The first assumption is confirmed only in part; indeed chemistry seems to follow a power law (although we cannot exclude other distributions) starting from a minimum value and this could be explained through a preferential attachment of scientific quality by researchers, namely researchers with a lower profile tend to tie to excellent researchers to increase the chances of funding. Laudel (2002) speaks about 'lobby' in project funding contexts, this effect can enhance the chances of funding success and could become a real 'lobby', with the risk of influencing research policy and decisions on the allocation of loans in its favour. Even if analyzing the presence of lobby is out of the scope of the paper, we can imagine that this kind of effect is also present in Italy and in other countries. Physics reveals a small-world characteristic (short paths and high cluster coefficient) (Watts and

**Table 7** SAOM for economics

| Parameters | Model 1 | | Model 2 (Final) | |
|---|---|---|---|---|
| | Estimate β | t-ratios | Estimate β | t-ratios |
| Rate (2000–2011) | 4.012 (average all period) | | 3.733 (average all period) | |
| Density | −2.252*** (0.229) | −0.164 | −1.655*** (0.993) | −0.194 |
| Transitive triplets | 1.101** (0.122) | 0.082 | 1.256** (0.428) | 0.026 |
| Betweenness | −0.511 (0.115) | −0.132 | −0.122 (0.986) | −0.698 |
| Degree assortativity | 0.364 (1.922) | 1.125 | | |
| Size university | | | 2.568 (1.336) | 2.328 |
| Collab (Call PRIN) | | | 2.339** (0.329) | 0.071 |
| Funding (Call PRIN) | | | 1.118** (0.355) | 0.054 |
| Gender (Call PRIN) | | | 1.371 (1.809) | 2.227 |
| Interdisciplinary | | | −2.137 (0.586) | −0.008 |
| Netchange (Call PRIN) | | | −1.320 (0.105) | −1.196 |
| Same academic role | | | 0.975 (0.337) | 0.023 |
| Gender | | | 0.115 (0.757) | 0.029 |
| Geographical proximity | | | −0.225** (0.029) | −0.006 |

"Estimation" = average of parameters' estimations. Standard deviation in parenthesis; "t-ratios" = test for the convergence of the algorithm (*t* values <0.1)

Significance of the estimation values (probability of acceptance of the null hypothesis): *** <0.01; ** <0.05; * <0.10

**Table 8** SAOM for sociology

| Parameters | Model 1 | | Model 2 (final) | |
|---|---|---|---|---|
| | Estimate β | t-ratios | Estimate β | t-ratios |
| Rate (2000–2011) | 4.956 (average all period) | | 4.236 (average all period) | |
| Density | −1.256** (0.079) | −0.029 | −1.655** (0.993) | −0.094 |
| Transitive triplets | 0.056** (0.236) | 0.033 | 2.256** (0.336) | 0.026 |
| Betweenness | −0.286 (0.786) | −0.028 | −0.322 (0.982) | −0.082 |
| Degree assortativity | −1.542 (0.034) | −0.076 | | |
| Size university | | | 1.904 (1.336) | −0.039 |
| Collab (Call PRIN) | | | 0.120 (0.329) | 0.001 |
| Funding (Call PRIN) | | | 0.034** (0.780) | 0.051 |
| Gender (Call PRIN) | | | 1.371 (0.029) | 0.006 |
| Interdisciplinary | | | −0.458(0.193)* | −0.001 |
| Netchange (Call PRIN) | | | −1.117 (1.028) | −0.059 |
| Same academic role | | | 0.018** (0.637) | 0.023 |
| Gender | | | 0.125 (0.957) | 0.001 |
| Geographical proximity | | | −0.803*** (0.568) | −0.026 |

"Estimation" = average of parameters' estimations. Standard deviation in parenthesis; "t-ratios" = test for the convergence of the algorithm (*t* values <0.1)

Significance of the estimation values (probability of acceptance of the null hypothesis): *** <0.01; ** <0.05; * <0.10

Strogatz [1998]) but does not exhibit a power-law degree distribution according to the Clauset et al. ([2009]) method. In physics some nodes are closer than the other fields, there is one large team that form the majority of the collaborations. This could be explained by the presence of large laboratories in physics, which require many researchers (therefore more connections). The cluster coefficient is present in all areas, this means that the friends of friends will have a larger probability to collaborate together. Altogether, however, the presented results are in agreement with Kronegger et al. ([2012]), supporting the conclusion that the dynamic social networks are governed by different forms between disciplines (in our case the network structure has different distribution for each area of study). Each field of study is affected by their own characteristics and academic culture. The second hypothesis is confirmed for all disciplinary fields. Data show that geographical proximity is an important driver for the longitudinal evolution of the network in all disciplinary fields. Even if we are in a national context (relative distance), physical proximity plays a key role in the formation of collaborations. It confirms the effects of face to face mentioned by Laudel ([2006]); for instance, researchers who work in the same department.

The third hypothesis is also confirmed; an interdisciplinary component does not play a role when a researcher chooses a partner of research. This does not mean that the projects are mostly disciplinary, but simply that researchers do not change the links on the base of partner's area. Finally, the last hypothesis about H-index is confirmed, there is no system of connection with those who have the same H-index. Therefore, there is a preferential attachment about the scientific prestige in chemistry and physics. Gender variable has been used as a control variable and it is not significant in the four models. This suggests that there are not mechanisms of gender homophily even if we cannot exclude an effect of interaction between gender and differences in rank (different academic role). This interaction could invalidate the effect on the dependent variable, because females are found less in powerful positions (e.g. positions of full professor) compared to males and so they might be less attractive. Moreover, research policies (isolated through the calls) drive the collaborative processes, as demonstrated by the significance of some control variables (e.g. Funding and Collab); this means that changes in the calls also influence the choice of research partners.

This paper concerns a specific Italian funding scheme in four areas of interest and the results are not valid for other funding schemes, in Italy and abroad; it cannot be generalized to the entire academic community.

## References

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics, 74*, 47.

Anand, B., & Khanna, T. (2000). Do firms learn to create value? The case of alliances. *Strategic Management Journal, 21*, 295–315.

Andersson, K. A. (2011). The formation of collaboration networks among individuals with heterogeneous skills. Working Papers. IDEAS: https://ideas.repec.org/p/cmu/gsiawp/938084531.html#biblio.

Andersson, A., & Persson, O. (1993). Networking scientists. *The Annals of Regional Science, 27*(1), 11–21.

Arenas, A., Danon, L., Díaz-Guilera, A., Gleiser, P., & Guimerá, R. (2004). Community analysis in social networks. *European Physical Journal B, 38*(2), 373.

Barabàsi, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509–512.

Beaver, D. D. (2001). Reflections on scientific collaboration (and its study): Past, present, and future. *Scientometrics, 52*, 365–377.

Bellotti, E. (2012). Getting funded. Multi-level network of physicists in Italy. *Social Networks, 34*, 215–229.

Bollen, J., Rodriguez, M. A., & de Sompel, H. V. (2006). Journal status. *Scientometrics, 69*(3), 669–687.

Bornmann, L., & Marx, W. (2011). The h index as a research performance indicator. *European Science Editing, 37*(3), 77–80.

Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies, 39*(1), 61–74.

Boschma, R. A., & Frenken, K. (2009). The spatial evolution of innovation networks. A proximity perspective. In R. A. Boschma & R. Martin (Eds.), *Handbook of evolutionary economic geography*. Cheltenham: Edward Elgar.

Chatterjee, A., Yarlagadda, S., & Chakrabarti, B. (2005). *Econophysics of wealth distributions*. Milan: Springer.

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review, 51*, 661–703.

Davis, P. M. (2008). Eigenfactor: Does the principle of repeated improvement result in better estimates than raw citation counts? *Journal of the American Society for Information Science and Technology, 59*(13), 2186–2188.

De Stefano, D., Fuccella, V., Prosperina, V. M., & Zaccarin, S. (2013). The use of different data sources in the analysis of co-authorship networks and scientific performance. *Social Networks, 35*(3), 370–381.

Durfee, E. H., Lesser, V. R., & Corkill, D. D. (1989). Cooperative distributed problem solving. In A. Barr, P. R. Cohen, & E. A. Feigenbaum (Eds.), *The handbook of artificial intelligence IV* (pp. 83–127). Reading, Massachusetts: Addison-Wesley.

Fernández, D. A., & López, J. M. V. (2015). Towards a European supranational policy of education based on the European dimension on education. *Bordón. Revista de Pedagogía, 67*(1), 101–116. doi:10.13042/Bordon.2015.67107.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. London: Sage.

Gillespie, C. S. (2015). Fitting heavy tailed distributions: The poweRlaw package. *Journal of Statistical Software*, *64*(2), 1–16. http://www.jstatsoft.org/v64/i02/.

Giot, L., et al. (2003). A protein interaction map of Drosophila melanogaster. *Science, 302*, 1727–1736.

Gmür, M. (2003). Co-citation analysis and the search for invisible colleges: A methodological evaluation. *Scientometrics, 57*(1), 27–57.

Hicks, D. M., & Katz, J. S. (1996). Where is science going? *Science, Technology and Human Values, 21*, 379–406.

Huisman, M., & Snijders, T. A. B. (2003). Statistical analysis of longitudinal network data with changing composition. *Sociological Methods & Research*, *32*(2), 253–287.

Huisman, M., & Steglich, C. E. G. (2008). Treatment of non-response in longitudinal network studies. *Social Networks, 30*, 297–308.

Hummon, N. P., & Carley, K. (1993). Social networks as normal science. *Social Networks, 15*, 71–106.

Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks, 11*, 39–63.

Katz, J. S. (1994). Geographical proximity and scientific collaboration. *Scientometrics, 31*(1), 31–43.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*, 1–18.

Kronegger, L., Mali, F., Ferligoj, A., & Doreian, P. (2012). Collaboration structures in Slovenian scientific communities. *Scientometrics, 90*(2), 631–647.

Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation, 11*, 3–15.

Laudel, G. (2006). *The 'quality myth': Promoting and hindering conditions for acquiring research funds*. Dordrecht: Springer.

Lepori, B., van den Besselaar, P., Dinges, M., Potì, B., Reale, E., Slipersaeter S., et al. (2007). Comparing the evolution of national research policies: What patterns of changes? *Science and Public Policy, 34*(6), 372–388.

Merton, R. K. (1968). The Matthew effect in science. *Science, 159*, 53–63.

Newman, M. E. J. (2001a). Clustering and preferential attachment in growing networks. *Physical Review E, 64*, 025102.

Newman, M. E. J. (2001b). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 404.

Newman, M. E. J. (2004). Who is the best connected scientist? A study of scientific co-authorship networks. In E. Ben-Naim, H. Frauenfelder, & Z. Toroczkai (Eds.), *Complex networks* (pp. 337–370). Berlin: Springer.

Newman, M. E. J. (2005). Power laws. Pareto distributions and Zipf's law. *Contemporary Physics, 46*, 323–351.

Nooteboom, B., Vanhaverbeke, W., Duysters, G., Gilsing, V., & Van Den Oord, A. (2007). Optimal cognitive distance and absorptive capacity. *Research Policy, 36*, 1016–1034.

Norris, J. (1997). *Markov chains*. Cambridge: Cambridge University Press.

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge: Cambridge University Press.

Reale, E, & Zinilli, A. (2014). Evaluation processes in research grant allocation: What is the relevance of scientific quality? In *Paper presented at the 27th CHER conference*. Rome. 8–10 September.

Robbins, H., & Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics, 22*, 400–407.

Scherngell, T., & Lata, R. (2011). Towards an integrated European research area? Findings from Eigen-vector spatially filtered spatial interaction models using European Framework Programme data. *Papers in Regional Science, 92*, 555–577.

Sinha, S. (2006). Evidence for power-law tail of the wealth distribution in India. *Physica A, 359*, 555–562.

Smith. D. & J. S. Katz (2000). Collaborative approaches to research. A report to the Higher Education Funding Council for England. Centre for Policy Studies in Education, University of Leeds.

Snijders, T. A. B. (1996). Stochastic actor-oriented models for network change. *The Journal of mathematical Sociology, 21*(1–2), 149–172.

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. In M. Sobel & M. Becker (Eds.), *Sociological methodology* (pp. 361–395). Boston, MA: Basil Blackwell.

Snijders, T. A. B. (2005). Models for longitudinal network data. In P. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 215–247). New York (NY): Cambridge University Press.

Snijders, T. A. B., Steglich, C., Schweinberger, M., & Huisman, M. (2010). *Manual for SIENA version 3.2. Provisional version*. Oxford: University of Oxford, Department of Statistics.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica, 57*, 307–333.

Watts, D. J., Dodds, P. S., & Newman, M. E. J. (2002). Identity and search in social networks. *Science, 296*, 1302–1305.

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

Zinilli, A., & Cerulli, G. (2015). DATANET: A stata procedure to facilitate dataset organization for network analysis, Statistical Software Components, Boston College Department of Economics. http://EconPapers.repec.org/RePEc:boc:bocode:s458033. Published: 06-25-2015.