Scientometrics (2012) 90:631–647
DOI 10.1007/s11192-011-0493-8

Collaboration structures in Slovenian scientific communities

Luka Kronegger · Franc Mali · Anuška Ferligoj · Patrick Doreian

Received: 6 June 2011 / Accepted: 25 August 2011 / Published online: 20 September 2011

Abstract We combine two seemingly distinct perspectives regarding the modeling of network dynamics. One perspective is found in the work of physicists and mathematicians who formally introduced the small world model and the mechanism of preferential attachment. The other perspective is sociological and focuses on the process of cumulative advantage and considers the agency of individual actors in a network. We test hypotheses, based on work drawn from these perspectives, regarding the structure and dynamics of scientific collaboration networks. The data we use are for four scientific disciplines in the Slovene system of science. The results deal with the overall topology of these networks and specific processes that generate them. The two perspectives can be joined to mutual benefit. Within this combined approach, the presence of small-world structures was confirmed. However preferential attachment is far more complex than advocates of a single autonomous mechanism claim.

Keywords Scientific collaboration · Co-authorship network · Bibliometry · Longitudinal network analysis · Small world · Preferential attachment · Stochastic actor based model

Introduction

Science has been studied from a variety of disciplinary vantage points and distinct methodologies have been used. Our primary goal is to couple new quantitative models

L. Kronegger (✉) · F. Mali · A. Ferligoj · P. Doreian
Faculty of Social Sciences, University of Ljubljana, Kardeljeva ploščad 5, 1000 Ljubljana, Slovenia
e-mail: luka.kronegger@fdv.uni-lj.si

F. Mali
e-mail: franc.mali@fdv.uni-lj.si

A. Ferligoj
e-mail: anuska.ferligoj@fdv.uni-lj.si

P. Doreian
Department of Sociology, University of Pittsburgh, 2602 WWPH, Pittsburgh, PA 15260, USA
e-mail: pitpat@pitt.edu

drawn from work by physicists and by social network analysts with information that is often viewed as more qualitative.[1] We use temporal data from the Slovene system of science that focuses on co-authorship networks in an effort to understand some of the dynamic aspects of scientific systems.

The early pioneering work of Price (1963, 1965), together with the research of Garfield summarized in Garfield (1979), set the stage for two important lines of activity. One was the creation of bibliometrics and scientometrics that permitted the study of large disciplinary systems. The other facilitated the location of individual scientists within these systems. The latter can be done in a general fashion but also to meet the need of scientists to know "what is going on" in the academic environments within which they work. The key feature permitting these lines of activity is the representation of scientific collaboration, conceptualized as network of scientists, and its operationalization in the form of co-authorship. Of course, there are other networks within the system of science but we focus specifically on co-authorship.

Since those early works of Price and of Garfield in the 1960s, sociologists introduced a series of theories that deal with scientific collaboration. Separately from these initial developments, a new scientific field emerged for examining social network analysis where mathematicians and physicists developed a range of methods that were later applied to networks of scientific collaboration. Here, we join these lines of inquiry by focusing on the theory of cumulative advantage in science, known also as the Matthew effect[2] (Merton 1968, 1973; Price 1976), and the theory of small world structure (de Sola Pool and Kochen 1978) and their applications to modeling of dynamics in co-authorship networks.

Translating the idea of cumulative advantage to research about the operation of science implies that those scientists who already occupy a position of excellence are rewarded far more than others in their field. Scientists who are rich in recognition find it easier to obtain additional recognition. In contrast, scientists who receive little recognition for their research efforts have reduced chances for future recognition.

Formal modeling of cumulative advantage in terms of preferential attachment was brought to social network analysis by Barabási and Albert (1999), who investigated a common property of many large networks whose vertex degrees follow a scale-free power-law distribution. This feature was found to be a consequence of two generic mechanisms: (i) networks expand continuously by the addition of new vertices, and (ii) new vertices attach preferentially to vertices that are already well connected. They presented a model based on these ingredients and reproduced the observed stationary scale-free distributions. Based on these results, they claimed that the development of large networks is governed by robust self-organizing phenomena that go beyond the particulars of the individual actors. The model was widely accepted and the implications of scale-free distributions were used to delineate the structure of scientific collaboration networks (e.g., Barabási et al. 2002; Moody 2004; Perc 2010; Kronegger et al. 2011). However, the notion of preferential

[1] In terms of substance, a rigid distinction between qualitative and quantitative approaches contributes little and focuses attention on a pointless division. Given that both approaches have merits and drawbacks, it seems more constructive to combine them to take advantage of their strengths.

[2] The idea of cumulative advantages comes from the passage in Matthew's Gospel: "For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath." Hence the term "the Matthew effect". The first systematic representation of cumulative advantage in science was provided by Merton (1973). Following him, a research stream invoked the idea of cumulative advantage as a central explanatory principle for the social stratification of science. Merton's studies were concerned with both organizational and functional aspects of science as an institution capable of self-regulation through scientists adopting a common set of norms about scientific conduct.

attachment reduces the generation of co-authorship to a single mechanism and ignores both institutional and contextual features of the environments within which scientists work.

The small-world network structure of scientific co-authorship implies network forms where the level of local clustering (one's collaborators are also collaborators with each other) is high and the average number of steps between clusters is small. In these small-world networks, internal ties of clusters tend to form and make the clusters of scientists more cohesive clusters. In contrast, ties between clusters are fewer and the network is less cohesive overall. However, paths between actors in different clusters tend to be short.[3]

The small-world model was formally defined by Watts and Strogatz (1998), who introduced an algorithm to construct networks with the following properties that mirror some observed social networks: i) having short paths between any two vertices (and hence, smaller average lengths for the shortest paths) and ii) also incorporates clustering (small dense parts of the network). These properties were later used to identify small-world structure in measured networks defined on co-authorship of scientific publications (e.g., Newman 2000, 2001; Moody 2004; Perc 2010). Perc examined the entire Slovene system of science for 1965 through 2010. He focused on the size of the largest component, a clustering coefficient, and mean distance between authors, all of which is consistent with the small-world model. He demonstrated that the network growth in Slovenia is exponential in time. Price (1963, 1965), taking an even longer term perspective, observed that the growth of networks in science followed the characteristic S-shaped logistic curve and speculated about there being three phases: (i) exponential (with sharp increases); (ii) linear (with a much slower rate of growth); and (iii) saturated (with a very slow rate of growth). The results of Perc suggest that science may still be in the first phase, at least in localized systems.

The formalizations of the preferential attachment and the small-world model were introduced by physicists who built the models primarily to reproduce the structure of real world networks at particular points in time. From the perspective of social scientists, these approaches ignore the social and institutional components of changes in networks through time. Science is not one homogeneous system but is comprised of many disciplines and sub-disciplines. If these disciplines change in different ways during different eras, this suggests that focusing on one mechanism (preferential attachment) and one network topology (small-world) provides a stunted view of scientific change. In particular, the notion of the small-world structure, may need to be qualified to take into account some of the social organization of science. The collaboration of scientists in closed groups can be influenced by additional features including interest in similar research topics (Kuhn 1996; Moody 2004). Other authors (e.g., Rodriguez and Pepe 2008; Ziman 1994) argue that co-authorship is primarily driven by departmental and institutional affiliation. An additional factor influencing collaboration is the mentor-student relationship described by Said et al. (2008) who indicated several styles of co-authorships. In contrast, the preferential attachment principle ignores these potential driving forces of collaboration by assuming

[3] According to various social network analysts, the small-world model was inspired by the work of de Sola Pool and Kochen (1978) who partially formalized the much more famous application of Travers and Milgram (1969). It expresses the simple idea that any two individuals, selected randomly from almost anywhere on the planet, are 'connected' via a path of no more than a small number of intermediate acquaintances. The (limited) empirical evidence suggested that this small number is about 6. This notion became a popular idea in the Broadway play named *Six Degrees of Separation*. The first practical evidence for the existence of a small-world phenomenon was first provided by the psychologist Milgram (Berg 2004, p. 46). Milgram's experimental result was regarded as a good starting point for analyzing the underlying structure of scientific co-authorship.

that individual scientists are motivated solely to co-author scientific documents with others having high (or higher) levels of scientific prestige.

In thinking about the applicability of preferential attachment to co-authorship it is necessary to be more specific about the types of relationships that are involved. For both links to websites from other websites and for citations between scientific productions, the notion of having high indegrees reflecting popularity and prestige respectively makes great sense. It costs little to add a link to another website or to cite another scientific production. Co-authorship, however, seems different because there are real time constraints on the creation of scientific productions. There is an element of circularity when the notion of preferential attachment is applied to co-authorship in the sense that: (i) the unequal distribution of co-authorship ties implies that this is driven by those with fewer co-authorship ties seeking to coauthor with those having more such ties; and (ii) having more collaborative ties is evidence of higher quality or prestige. We mean this in the sense of there being conceptual circularity. For the second part, it appears that a paper is a paper and all that is needed is a count of the number of coauthored publications. In using the term circularity we do not mean to imply that there is not an endogenous process whereby quality is an emergent phenomenon. But this is far more likely to occur in a citation network than in a co-authorship network. Of course, when some scientists are known to produce quality work, or to have produced quality work, there can be an incentive for others to try and collaborate with them. It is not clear that there is an incentive for prestigious scientists to work with others, especially if they are unknown to them. This immediately raises the issue of access because collaboration, at a minimum, requires access and this is more likely to be a more potent predictor of collaboration than counts of coauthored productions.

In support of the idea that collaboration is driven, in part, by access we argue that access is facilitated greatly by working at the same location as well as sharing an interest in the same topic(s). For example, if a prominent scientist runs a large funded program, this is likely to lead to more scientific productions and more co-authorship ties. A chain of causation is, roughly, something like funding → organizational infrastructure → joint work → collaborative ties within the infrastructure. Even if they do not belong to an organized research program, two scientists at the same location have an increased chance of collaborating by virtue of working in the same place. Collaboration in general, and co-authorship in particular, can take various forms: within disciplines, across disciplines and across national boundaries. Evidence from earlier research provide support for a positive effects of cross-national collaboration on the quality of research performance of scientists. See, for example, Abramo et al. (2011); Glänzel and de Lange (2002). Included among the indicators used in such studies is a measure of the quality of journals. Admittedly, measures such as journal impact factors are crude but they do provide separate indicators of quality. We include these other potential variables and test their relevance when preferential attachment is included (and vice versa)[4]. We operationalize preferential attachment as the number of different collaborators (degree) of a single researcher while including the number of good publications etc. (e.g., Hara et al. 2003; Börner et al. 2005).

Until recently, no methods were available for modeling change in dynamic networks while including actor attributes and organizational contexts as factors influencing change in networks. With the development of stochastic actor-oriented models this has changed and it is now possible to estimate complex models consisting of a series of interacting

[4] We included the following variables: the number of coauthors within the national border of the discipline and the number of coauthors coming from outside of the discipline.

micro-mechanisms that drive change in networks. These models can be tested by using temporal network data together with potential intervening variables and evaluate their performance through statistical inference. These methods for modeling network dynamics have been implemented in a variety of packages. Here, we use SIENA (Snijders et al. 2008, 2010). These techniques permit a comparative assessment of small-world models, preferential attachment, and factors capturing actor agency and organizational contexts.

Specification of model(s)

The small-world model of Watts and Strogatz (1998) has two well defined characteristics: (i) clustering; and (ii) short path lengths. Its point of departure of having nodes on a regular lattice is followed by random rewiring of ties and is rather artificial. Yet the final structure has noteworthy features that are apparent in many real world networks if we allow that the mechanisms for this topology are likely to be different. Instead of random rewiring there is actor agency regarding the creation and deletion of ties rather than having the implicit mechanism residing in the ties themselves (Robins et al. 2005). They show also that small-world models are fully compatible with ERGMs, in the form of stochastic actor-oriented models, where the micro-mechanisms are found in some local structural configurations of ties. We follow this approach here by incorporating both the small-world ideas and preferential attachment in an ERGM together with other factors that can drive collaboration.

Small-world structure

The basic hypothesis regarding the operation of small-world processes is simple to state in a single statement that can be separated into two distinct parts:

H1 The co-authorship networks for science for all disciplines have a small-world structure.

H1a The topology of co-authorship networks for science has separated dense patches of ties.

H1b Paths between pairs of scientists in the co-authorship networks are short.

As noted earlier, the small-world model developed by Watts and Strogatz (1998), has two important operationalized characteristics. Such networks have high values of the clustering coefficients. Intuitively, a clustering coefficient represents the average probability that two neighbors of one author also collaborate. More precisely, clustering coefficient is defined as follows:

$$C_n = \frac{2e_n}{(k_n(k_n - 1))}$$

where k_n is the number of neighbors of the author n and e_n is the number of connected pairs between all neighbors of the author n (Watts and Strogatz 1998).

When a coefficient for an actual network is computed it has to be evaluated by comparing the value to an expected value. The latter can be calculated theoretically, or estimated from simulated random networks having the same characteristics of an actual network using Erdős–Rényi model. A high clustering coefficient, as a global characteristic of the network, can be viewed also as a consequence of transitivity. This captures the activity of scientists on local level. The tendency of actors to form transitive ties is included in our model.

The second characteristic of the small-world model, short average shortest path between two vertices, is hard to operationalize in terms of driving factors. The average path in the network is a quantity that has no direct connection to the behavior of individual actors in the network. While actors can track short paths of ties, there is little evidence to suggest that actors can, or do, track long paths. So this is not included as a potential driving factor. Short average path length is a consequence of the operation of other processes and so it cannot be ignored even though it is not included as a variable. Information on the average path within the co-authorship networks of disciplines, together with the expected average path length of random networks having the same broad features of the observed networks, is presented separately in Table 1 and discussed when interpreting the estimated models.

As discussed by Moody (2004), a major mechanism encouraging the formation of small-world structures in scientific collaboration networks is the fragmentation of disciplines to sub-disciplines and specific topics. One variable capturing this is research group membership, in relation to scientific collaboration (Rodriguez and Pepe 2008). This is especially relevant in the Slovene context and captures institutional affiliations and geographic proximity of the actors. The organization of research groups in Slovene scientific system corresponds to some general and in most cases unique research topics.[5] Therefore the variable can also be used as proxy for 'research topic similarity', an operationalization of which is not a part of our database. In presenting our results, we use 'the same research group' to represent this variable.

The second variable used as a driver of collaboration among scientists is the year of the first publication (for each scientist) within the database. We use this variable as a proxy for the scientific age of a researcher. With age and age similarity we can therefore test the strength and role of seniority, or mentorship, for young-researcher or student ties within the structure formation of scientific collaboration network.

Preferential attachment

While our first hypothesis focuses on the overall structure of co-authorship networks, the second hypothesis deals with a mechanism for the structure of these networks.

H2 The structure of the co-authorship networks is driven by preferential attachment mechanism.

Preferential attachment (Barabási and Albert 1999) as a model, shifts from modeling network topology to modeling network assembly and evolution. This model is consistent with a production process that Moody (2004) calls 'star production' in which authors with many collaborators and high scientific prestige obtain more attention and connections from authors that are joining the network than the other scientists. The key property of networks growing in accordance preferential attachment is that their degree distribution follows a power-law, at least asymptotically. That is, the fraction $P(k)$–$k^{-\gamma}$ of nodes in the network

[5] In Slovenia, the organization of research group around the unique research topics is (artificially) encouraged by the use of some R&D policy instruments. One very important instrument of governmental R&D policy are research programs at public universities and institutes because they are financed by the Slovene Research Agency. These research programs cover research topics that are financed for very long periods (up to seven or more years). Such long-term financial stability of research programs reduces the flexibility of research topics. One consequence takes the form of rigid and closed research groups.

Table 1 Network properties through time

	t_1 1991–1995	t_2 1996–2000	t_3 2001–2005	$t_{1\text{-}3}$ 1991–2005
Physics				
Number of vertices	125	183	234	245
Number of edges	274	487	686	938
Average degree	4.38	5.32	5.86	7.66
Average distance	3.44	4.8	5.15	3.97
Expected distance[a]	3.75	4.20	3.10	3.75
Clustering coefficient	0.461	0.473	0.492	0.437
Expected CC[a]	0.032	0.031	0.014	0.032
Mathematics				
Number of vertices	65	96	135	142
Number of edges	42	63	122	157
Average degree	1.29	1.30	1.81	2.21
Average distance	2.34	3.94	4.52	4.36
Expected distance[a]	3.41	4.16	3.45	3.41
Clustering coefficient	0.246	0.302	0.285	0.254
Expected CC[a]	0.047	0.044	0.073	0.046
Biotechnology				
Number of vertices	33	50	79	86
Number of edges	42	58	147	180
Average degree	2.47	2.32	3.72	4.19
Average distance	2.45	2.88	3.34	3.91
Expected distance[a]	6.62	5.39	4.36	6.63
Clustering coefficient	0.555	0.339	0.480	0.440
Expected CC[a]	0.013	0.013	0.018	0.013
Sociology				
Number of vertices	61	88	111	114
Number of edges	26	124	199	253
Average degree	0.85	2.82	3.59	4.44
Average distance	1.74	3.14	3.37	3.00
Expected distance[a]	3.11	3.08	3.14	3.28
Clustering coefficient	0.500	0.589	0.539	0.478
Expected CC[a]	0.028	0.033	0.040	0.025

[a] Expected values are calculated on Erdős–Rényi graphs on given number of vertices and average degree with 10,000 repetitions

having k connections to other nodes goes for large values of k where γ is a constant whose value is typically in the range $2 < \gamma < 3$. In practice, when these distributions are plotted on log–log scale, they fit a straight line. These plots provide one way of testing **H2**.

The problem of actor agency remains problematic for all approaches to studying collaboration. It seems clear that scientists who collaborate and/or co-author scientific productions choose to do so. In short, they act and agency is implicitly part of the preferential attachment mechanism implying agency of some sort: scientists choose to coauthor with scientists having higher degrees in the co-authorship network. This is viewed as a micro-level mechanism involving actor agency having macro-structural consequences. That is, when all scientists behave in this fashion, these separate actions result in a network with, among other things, a co-authorship degree distribution that is a power law. However, there are multiple potential mechanisms that are ignored or left implicit. We have suggested that involvement in a common large project can generate co-authorships. If anything, this is—or includes—higher prestige scientists incorporating lower prestige scientists in some joint productions. A junior colleague can produce a manuscript to which the project leader then adds her/his name to it. Similarly, scientists of higher prestige can add their names to help the chances of junior colleagues, or protégés, having a first author publication. Are these examples of preferential attachment? In general, they are not and the more general problem is that these differences in how a particular collaboration/co-authorship arrangement formed are lost in the preferential attachment formulation. Actually, this can be extended to many, if not all, potential micro-mechanisms with preferential attachment being just one of them. Unless there is direct evidence regarding the origins of a collaboration it seems that evidence for agency will always be implicit. We attempt to deal with this in our third hypothesis (below).

Another way of testing this proposition stems from think about how co-authorship networks are held together if preferential attachment is operative. Such networks must be held together by those prominent scientists having the highest degrees. It follows that if we exclude the most connected individuals in such network, it must break up into unconnected components (in the graph theoretical sense). This is measured with a component sensitivity indicator, which is included into the plots of the degree distributions. The component sensitivity is the number of components in the network after we exclude the researchers with highest degrees.

When we observe the preferential attachment principle, through the eyes of single unit, it can be modeled simply as the effect of the degree parameter on production of new ties.

Actor agency and organizational contexts

As noted above, preferential attachment is conceived as a mechanism that operates at the level of whole network. The small-world topology also concerns the structure of the whole co-authorship network. Of course, it is straightforward to argue that this applies also to specific coherent parts of the network. Vertex degree captures one feature resulting from the operation preferential attachment and the clustering coefficients are used to describe a feature of the network topology. However, underlying our discussion of the small-world model and preferential attachment there are other variables that can be viewed as affecting the evolution of scientific co-authorship networks. They can be subsumed into a broad hypothesis:

H3 Organizational and institutional contexts drive the formation of scientific co-authorship networks.

Scientists working in the same research group reflect one feature of the local organization of science and we include it as a variable. Within the ERGM approach, local structural configurations are included as predictors. One of them is transitivity in the form of completed triples is included as another mechanism for generating co-authorship ties. Of course, transitivity can be a local mechanism within organizations and also as a part of a wider network that extends beyond local organizational or institutional contexts. Even so, it is a structural feature not considered within the rubric of preferential attachment.

National scientific systems are organized within national states with both funding and social organizational mechanisms. If follows that this institutional context cannot be ignored. The co-authorship networks we study are defined for complete disciplinary systems within Slovenia. However, as noted earlier, scientific collaboration extends across both disciplinary boundaries and national borders. We anticipate that both, when operationalized, will be predictive of tie formation among scientists. Finally, even though it is an imperfect measure of quality, impact factors (IFs) are used, and promoted by publishers, as a measure of quality. Considering the IF of scientific journals is one way of getting around the conceptual circularity noted earlier by treating the IF as a more direct measure of quality.

Data

Our data set was obtained from two commonly connected sources in Slovenia: (i) the Current Research Information System (SICRIS) which includes the information on all active researchers registered at the Slovenian Research Agency and (ii) the Co-operative On-Line Bibliographic System & Services (COBISS) which contains a database of all publications that can be located through Slovenian libraries[6] Connecting these systems gives a unique officially maintained database of complete personal bibliographies of all researchers registered in Slovenia. SICRIS provides additional information on the education, positions and employment of researchers, information on the research groups and the institutions as well as information on both the projects and programs involving Slovenian researchers. Both systems are maintained by the Institute of Information Science in Maribor (IZUM).

For our analysis we selected all researchers who were registered as working in four disciplines: physics; mathematics; biotechnology; and sociology[7] and were in 2008 included in the SICRIS database. The selection of disciplines was guided by prior research of co-authorship networks (Newman 2004; Barabási et al. 2002; Moody 2004) and interpreted through Hargens' model of functional and normative integration (Hargens 1975). The selection covers different types of disciplines and disciplinary cultures. Physics can be viewed as an old, well established discipline with scientific processes organized within research teams and laboratories. Biotechnology has similar organization of work (with research teams and laboratories) but is a young discipline that is still being established. Mathematics is a so-called office discipline, where collaboration, when there is any, does not take place in a laboratory but involves solving globally abstract scientific problems. Sociology shares some similarities to mathematics as an office discipline but, as part of social sciences, is more focused on collaboration in small groups dealing with local issues.

We obtained information for all scientists on their scientific publications published in 1991 through 2005 and generated networks for each discipline according to co-authorship of publications. Scientific publications are selected by IZUM according to criteria for evaluations of scientific performance used in the Slovenian Research Agency, This

[6] This does not imply that they are actually in a library, although most of them are. Researchers are required to document scientific production to a senior scientific librarian who verifies these documents and includes them in a listing of scientific documents.

[7] In a 'general typology' of science in Slovenia, physics, mathematics and sociology are root (first level) disciplines while biotechnology belongs to biotechnological sciences and is therefore second-level category.

selection is intended to be inclusive: all scientific articles, monographs, chapters in monographs, published scientific conference contributions, patents, scientific databases or corpuses and scientific films, sound or video recordings. For the analyses presented here, the boundaries of the disciplinary networks are defined within the SICRIS database. All connections of researchers to authors outside the selected disciplines within Slovenia were summed for each researcher and operationalize a variable indicating the extent of collaboration outside the discipline. Similarly, the sum co-authored publications with scientists outside Slovenia operationalizes external co-authorship. For the purpose of estimating stochastic actor-oriented models, networks were constructed for three time slices as co-authorship networks for three periods: 1991–1995, 1996–2000 and 2001–2005. Some basic information on these networks is presented in Table 1, together with information for the entire period 1991–2005. Detailed analysis of these data can be found in Kronegger et al. (2011).

Results

We start the detailed analysis using small world and preferential attachment model and continue with stochastic actor based modeling. At the end of this section we compare both approaches, describe similarities and examine reasons for the differences we identify.

Modeling real world networks

The key global indicators of small-world networks are average distances between scientists in the network and clustering coefficients. Both are compared to what would be expected from the corresponding random networks. For physics, mathematics and biotechnology, the average distances are all shorter than what would be expected in a corresponding random network. The reverse is the case for sociology (except for t_2) with values that are close to what would be expected from a corresponding random network. For all four disciplinary networks, the average distance is increasing through time, consistent with the increasing size of all these networks. The average distances in the networks through all three time periods (t_{1-3}), are 3.97 for physics, 4.36 for biotechnology, 3.91 for mathematics and 3.00 in sociology.

It is known that clustering coefficients of real world networks, in most cases, far exceed the levels of these coefficients for corresponding random networks. Our results are fully consistent with this for all networks and all time periods. The clustering coefficient (for t_{1-3}) is about 0.48 for sociology, 0.44 for biotechnology and physics, and 0.25 for mathematics. We can compare directly our results for sociology with those of Moody (2004) who analyzed networks of sociologists generated from articles published in *Sociological Abstracts*. He reports clustering coefficient between 0.2 an 0.3. One reason for our higher clustering coefficients is probably the consequence of the definition of our network boundaries. Our networks are for whole scientific communities within one specific small country. In addition to having smaller networks, they are bounded by national borders and organizational structure of science on the national level. Both raise the probability of forming transitive ties. Moody also studied publications in biology, physics and mathematics, reported clustering coefficients of: 0.07 for biology; 0.36 for physics; and 0.12 for mathematics.

Perc (2010) analyzed the co-authorship network of all Slovenian scientists (which was generated from the same data source as ours) and reported a linearly decreasing clustering coefficient from 0.35 in 1990 to 0.23 in 2005. In contrast, the clustering coefficients increase through the three time periods we study for physics.[8] We note that Perc considered only scientific papers which is in accordance with most bibliometric analyses. For the other three disciplines these coefficients fluctuate, albeit in different ways, with no evident correspondence to the falling trend found by Perc. Because co-authorship leads to other types of scientific productions including monographs, patents, book chapters and corpuses, our database is more general. Given the variation across the time points and fields in our data, it is likely that studies of the whole network of science in Slovenia (or for any such system) bury variations within fields.

Even so, the global properties of co-authorship networks of physicists, mathematicians, biotechnologists, and sociologists indicate that their networks are formed according to the small-world principle. The key properties are relatively stable with the small changes mainly due to the growth of the networks. If anything, the small-world structure may be magnified for complete scientific systems of individual nations.

We turn now to consider the preferential attachment principle. The main indicator of a structure induced by preferential attachment structure is a scale-free distribution of degrees.[9] Fits of the scale-free functions of the degree distributions presented in Fig. 1, in the form of log–log plots, reveal no obvious clear preferential attachment mechanism guiding the development of the networks for both physics and biotechnology. To the extent that confirmation of preferential attachment is sought in the form of a power-law holding, this is a reasonable conclusion. However, if preferential attachment is but one of several mechanisms, it is possible that a power-law can be only approximated even though preferential attachment might be operative. It follows that more precision is needed regarding the conditions under which claims regarding preferential attachment are supported or refuted. At a minimum, our results suggest that if preferential attachment is operative then it is but one part of a set of mechanisms under which co-authorship ties are formed. The straight line fits well for both mathematics and sociology. This can be viewed as surprising to the extent that these are office disciplines rather than lab disciplines.

Another indicator of preferential attachment driven structure within the network is low stability of the network regarding the number of components when we exclude the most connected units. At face value, the star scientists, defined as those collaborating with many others, occupy an important part of the co-authorship network. If they have many co-authorship links, this suggests that their removal from the network would have a dramatic impact on the remaining network. In contrast, removing authors who collaborate seldom would have little impact. More precisely, removing these high degree scientists would diminish the extent to which the remaining networks are connected. If the network is held together by the most prominent—the star scientists—it should fragment into components after we remove them from the network. If the largest component does not dissolve after we remove the most connected scientists, we can question the hypothesis of a preferential attachment structure. The stability is shown in the diagrams as component sensitivity. The most stable network of the four we analyzed is the network of physicists. It is clear from the figure, that the networks hold together well until we remove all the 35 authors with at least 12 co-authors in physics. The corresponding figures for the other three disciplines is:

[8] Newman (2001) found that the physical sciences have much higher clustering coefficients than biomedicine. He concluded that one reason for this is the ‘top down’ organization of laboratories in classical physics. In contrast, within biomedicine it is less common for two scientists to collaborate if they have another collaborator in common.

[9] We exclude all scientists with degree zero for these plots.

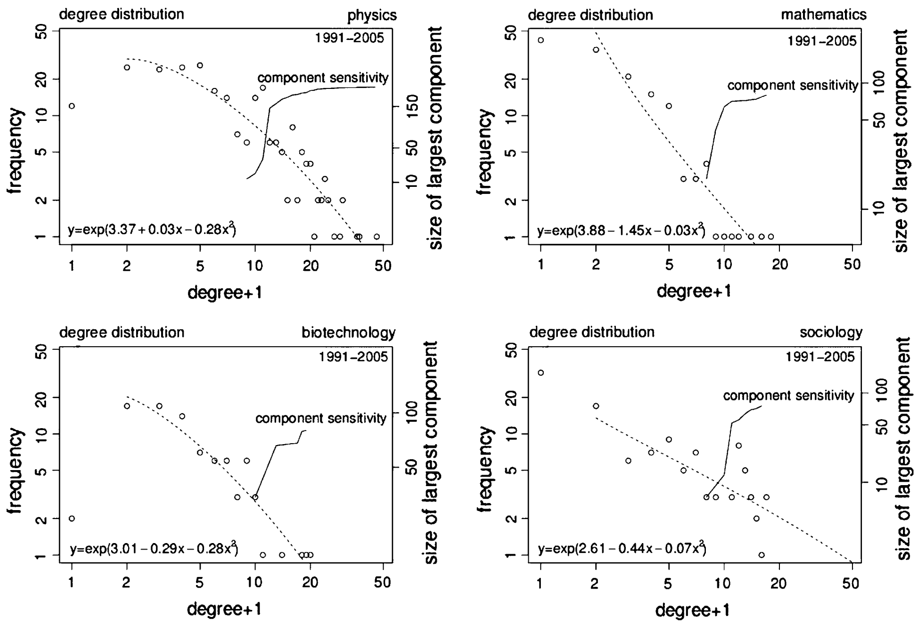

Fig. 1 Degree distribution and component sensitivity within so-authorship networks

5 authors with at least 10 so-authors in mathematics; four authors with at least 13 co-authors in biotechnology; and 22 authors with at least 11 co-authors in sociology. This suggests that the co-authorship networks in Slovenia for mathematics and biotechnology are held together by very small number of important authors. This does not appear to be the case for physics and sociology.

Stochastic actor-oriented model results

With use of stochastic actor based modeling we move from a bird's eye perspective of the overall (macro-level) networks to a micro-level view for observing co-authorship networks. Given the formulation underlying the construction of the SIENA program, it is possible to incorporate some organizational and institutional features. Implicitly, the preferential mechanism focuses solely on the formation of co-authorship ties. Yet, scientists collaborating at one point in time need not do so at a later time point. Or, if they do so, there can be a considerable interval between collaborative ventures. One feature of the ERGMs implemented in SIENA is that ties can be both formed and dissolved and, as this is a feature that characterizes collaboration and co-authorship networks, using these models is appropriate.

The dynamic modeling of networks requires network measurement for at least two measurements in time. In our case, we have networks measured in three waves, labeled t_1, t_2 and t_3, which span the years 1991–2005.

First, we consider the three parameters in the first panel of Table 2: the rate parameter for the first transition; the rate parameter for the second transition and the density parameter. They are all part of the overall model and are included for technical assumptions of the model. They are not especially crucial for the interpretation of the fitted

Table 2 Fitted model parameters for four scientific disciplines

	Parameters	Physics		Mathematics		Biotechnology		Sociology	
		Param.	Err.	Param.	Err.	Param.	Err.	Param.	Err.
1.	Rate 1	13.28[a]	(1.35)	2.37[a]	(0.46)	4.77[a]	(1.42)	21.19[a]	(4.67)
2.	Rate 2	11.03[a]	(0.89)	4.46[a]	(0.65)	4.88[a]	(0.79)	9.12[a]	(1.29)
3.	Degree (density)	−1.12[a]	(0.10)	−2.16[a]	(0.23)	−1.84[a]	(0.38)	−1.92[a]	(0.17)
4.	Transitive triads	0.47[a]	(0.03)	0.97[a]	(0.14)	1.11[a]	(0.14)	0.58[a]	(0.07)
5.	Same res. group	0.84[a]	(0.06)	0.39[a]	(0.11)	1.13[a]	(0.16)	1.28[a]	(0.10)
6.	First publ. similarity	−0.30[a]	(0.15)	0.10	(0.27)	−1.28[a]	(0.35)	−0.08	(0.22)
7.	Collaboration within disc. (sqrt)	−0.22[a]	(0.05)	0.25	(0.16)	0.07	(0.24)	0.22[a]	(0.09)
8.	Collaboration outside disc. (sqrt)	0.00[a]	(0.00)	0.00	(0.00)	−0.00	(0.00)	−0.00[a]	(0.00)
9.	No. of articles with IF	−0.00	(0.00)	0.02[a]	(0.01)	0.07	(0.04)	0.12[a]	(0.06)
10.	Year of First publ.	0.01[a]	(0.00)	−0.01	(0.01)	0.01	(0.02)	−0.01	(0.01)

The parameters significant at the 5% significance level are denoted by [a]

models. From the two rate parameters we can get the estimated average frequency of unobserved changes (per actor) within the networks. There are: 13 such changes (tie formation or dissolution) in the network of physicists in the transition of network from t_1 to t_2; 21 in the network of sociologists; 4.8 in the network of biotechnologists; and 2.4 in the network of mathematicians. For the next transition, the average number of changes for single units in the network rises slightly in the networks of mathematicians and biotechnologists while the estimated number of unobserved changes falls in the networks of physicists and sociologists.

Values of the third basic parameter (degree) are negative for all four disciplinary networks, which is an expected consequence of the costs of each tie formation for each researcher. The conception that tie formation has costs in terms of time, effort, and resources is important because researchers can co-author with only a limited number of different authors and each new tie presents a certain burden. The creation of some new ties, given the costs of tie formation, can lead also to the dissolution of other ties. Much of the discussion about preferential attachment is silent about the costs of ties in terms of tie creation and maintenance and appears to implicitly assume these acts are without cost.

We turn to consider the parameters used to model small world processes for the disciplinary networks. They are shown in Table 2. The fourth parameter of the model, transitive triads, captures the tendency of actors to form a ties in such a fashion that they close the triangles. Recall, if significant, it shows that scientists tend to form a new co-authorship tie with co-authors of their co-authors. The parameter is positive and significant for all four disciplines. This is undeniable confirmation of high clustering within the networks, a clear indicator of small-world network structures for these disciplines.

The value of the fifth parameter (second in this panel) represents the impact of belonging to the same research group parameter as a predictor of the tendency to form a new co-authorship tie. The parameters are positive and significant for all four disciplines which is an indicator of strong influence of formal institutional structures on scientific collaboration.

The sixth parameter also describes an external effect on the network dynamics. It is defined as the first publication similarity, which is used to test the tendency of researchers who have similar scientific experiences, or the same scientific age, to form a new a co-authorship tie. The value of this parameter is negative and significant only for the network of biotechnologists which indicates that ties are formed between researchers with different scientific ages (e.g., students and mentors). There is no such effect for physics, mathematics, and sociology.

In summary, the small-world network structure found in the previous section has been confirmed as operating in all four disciplines studied here. Additionally, we claim tie formation is governed also by research group membership in all four disciplines. Put differently, there is no *single* mechanism for co-authorship tie creation. Finally, for biotechnology the clustering mechanism is also driven by the collaboration between researchers with different levels of scientific experience.

We return again to the preferential attachment principle and consider the parameters in the bottom of Table 2. From the level of individual researcher the preference to make ties with researchers already having a high number of co-authors is clear. Because of the skewness of the distributions of the degree of collaboration we used the square root of the degrees in the model. However, the high degree of prominent authors is the result of a complex combination of factors. We include only the most obvious properties that could stimulate researchers to form new ties to researchers having specific characteristics. The variables we include are: (i) the square root of the degree within the network, which indicates number of co-authors within the national borders of the discipline, implemented as endogenous degree effect; (ii) the square root of the degree coming from publications outside the discipline; (iii) the number of articles published within journals having a recorded impact factor (regardless of its size); and (iv) year of the first publication. All appear to be plausible indicators of preferential attachment.

The number of collaborators within the network, is statistically significant in the networks of physicists and sociologists. The estimated parameter in physics is negatively signed which indicates, that those scientists who collaborate more within the discipline do not tend to form new ties with researchers from within the discipline. The opposite is true in sociology where this effect is positive.

When we look at the number of co-authors outside the discipline, the parameters having significant values are for physicists and sociologists. For physicists, the estimated parameter is positive which means that those scientists who collaborate with many authors from other disciplines (or from abroad) are more likely to form new ties with scientists from the discipline. For sociologists, the sign of this significant parameter is negative: among sociologists in Slovenia, collaboration with other researchers (outside the field or Slovenia) has a negative effect on tie formation with scientists working in sociology.

The number of articles published in journals having a recorded impact factor, matters in mathematics and sociology. Those mathematicians and sociologists who publish more in good (or higher ranked) journals tend to form new co-authorship ties more often with other researchers from their disciplines. While sociologists in Slovenia publish less than 5% of their scientific publications in journals having an impact factor, the percentage of researchers from all three natural disciplines exceeds 30% of such publications (Kronegger et al. 2011). This makes publishing of articles in the journals with an impact factor a much higher level of prestige boost among Slovene sociologists than among researchers from other analyzed disciplines in Slovenia.

Finally, the year of the first publication has positive significant effect only in physics. Those who are scientifically younger are more likely to establish co-authorship with other Slovenian physicists.

The results of fitting a stochastic actor-oriented model lead us to conclude that researchers in all four disciplines do form new co-authorship ties in a way that is consistent with the small-world structure of networks. At the same time, the mechanism of preferential attachment is far more complex than advocates of a global autonomous mechanism claim. To the extent to which it operates, which we do not dispute, it is connected in subtle ways to the social organization and unique collaboration culture of each discipline:

– In physics those who collaborate with high number of researchers from the discipline have lower probabilities of establishing new ties with other physicists within the Slovenian research community. This can be the consequence of the saturation of the network where scientific work is organized within the formalized environments around expensive and complicated technical research equipment (e.g., laboratories). Physicists that have collaborated extensively outside the Slovenian physics field have higher probability of establishing ties within the national physics community.
– The situation is different among mathematicians. Those who already collaborate extensively outside the Slovenian mathematical research community and have larger number of articles published in journals with impact factor have higher probabilities of establishing new ties with the researchers inside that research community.
– None of the parameters used to model the preferential attachment are significant in biotechnology, the youngest of the analyzed disciplines. This suggests that the dynamics driving co-authorship comes from other mechanisms. One such mechanism is a high level of clustering—just like in other disciplines. Another is membership in the same research groups and also differences in the scientific experience of the researchers. The latter may be a consequence of tie formation between mentors and their students.
– In sociology, extensive collaboration within the Slovenian research community has positive effect and collaboration outside that research community has negative consequences on collaboration within the national research community. At the same time, where the number of articles having an impact factor has a positive effect on the formation of new co-authorship relationships within the discipline.

Conclusion

We have coupled two seemingly different approaches to the study of scientific co-authorship networks and used them to formulate hypotheses about the topology of these networks and the mechanisms driving the formation of ties and hence the networks. We presented descriptive information about the co-authorship networks of four disciplines within the Slovene scientific system. More importantly, we have modeled some of the mechanisms that drive the formation of co-authorship networks for these disciplines. The first hypothesis about the presence of a small-world structure has been confirmed unequivocally by using both of the two approaches. The evidence regarding the second hypothesis concerning preferential attachment as the driving mechanism of co-authorship is, at best, mixed. Some features of this principle were confirmed for mathematics and sociology but not for physics and biotechnology. The third hypothesis, while overly broad, was confirmed and the evidence demonstrates that the four disciplines are affected, albeit in different ways, by the organization of local institutions and publishing cultures.

Some important implications follow from the results that we report. One is that it is more sensible to couple different perspectives regarding the mechanics of scientific co-authorships rather than hold them apart. The perspective created largely by physicists can be joined with the older sociological perspective to mutual benefit. A second implication is that, while science does operate in terms of general norms regarding participation, the realizations of these processes take different forms in different national systems. Local institutional arrangements matter. A fourth implication is that science is not one general phenomenon and the differences between disciplines also matter. Fifth, the processes of co-authorship, even for a single discipline, operate differently in different eras. Finally, by combining perspectives, mobilizing a coherent methodology, considering different disciplines (and different mixes of disciplines), and looking at different eras permits a more nuanced image of science and scientific processes.

Acknowledgments We are very grateful to the Institute of Information Science (IZUM), for the preparation of the datasets, to Tom Snijders and to the anonymous referees for their constructive comments.

References

Abramo, G., D'Angela, C.A., & Solazzi, M. (2011). The relationship between scientist's research performance and the degree of internationalization of their research. *Scientometrics*, *86*(3), 629–643.

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

Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, *311*(3–4), 590–614.

Berg, C. (2004). *Vernetzung als Syndrom. Risiken und Chancen von Vrenetzungsprozessen fuer eine nachhaltige Entwicklung*. Frankfurt: Campus Verlag.

Börner, K., Dall'Asta, L., Ke, W., & Vespignani, A. (2005). Studying the emerging global brain: Analysing and visualizing the impact of co-autorship teams. *Complexity*, *10*(4), 57–67.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, *1*(4), 359–375.

Glänzel, W., & de Lange, C. (2002). A distributional approach to multinationality measures of international scientific collaboration. *Scientometrics*, *54*(1), 75–89. doi:10.1023/A:1015684505035.

Hara, N., Solomon, P., Kim, S. L., & Sonnenwald, D. H. (2003). An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. *Journal of the American Society for Information Science and Technology*, *54*(10), 952–965. doi:10.1002/asi.10291.

Hargens, L. L. (1975). *Patterns of scientific research*. Washington, DC: American Sociological Association.

Kronegger, L., Ferligoj, A., & Doreian, P. (2011). On the dynamics of national scientific systems. *Quality & Quantity*, *45*(5), 989–1015. doi:10.1007/s11135-011-9484-3.

Kuhn, T. S. (1996). *The structure of scientific revolutions, 3rd edn*. Chicago: University Of Chicago Press.

Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*, 56–63.

Merton, R. K. (1973). *Sociology of science*. Chicago: Chicago University Press.

Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, *69*(2), 213–238.

Newman, M. E. J. (2000). *Small worlds: The structure of social networks*. Santa Fe: Santa Fe Institute.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 404–409. doi:10.1073/pnas.021544898.

Newman, M. E. J. (2004). Co-authorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(Suppl 1), 5200–5205.

Perc, M. (2010). Growth and structure of Slovenia's scientific collaboration network. *Journal of Informetrics*, *4*, 475–482. doi:10.1016/j.joi.2010.04.003.

Price, D. S. (1963). *Little science, big science and beyond*. New York: Columbia University Press.

Price, D. S. (1965). Networks of scientific papers. *Science*, *149*, 510–515.

Price, D. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, *27*(5), 292–306. doi:10.1002/asi.4630270505.

Robins, G. L., Woolcock, J., & Pattison, P. (2005). Small and other worlds: Global network structures from local processes. *American Journal of Sociology*, *110*, 894–936.

Rodriguez, M. A., & Pepe, A. (2008). On the relationship between the structural and socioacademic communities of a co-authorship network. *Journal of Informetrics*, *2*(3), 195–201. doi:10.1016/j.joi.2008.04.002.

Said, Y. H., Wegman, E. J., Sharabati, W. K., & Rigsby, J. (2008). Social networks of author–coauthor relationships. *Computational Statistics & Data Analysis*, *52*(4), 2177–2184. doi:10.1016/j.csda.2007.07.021.

Snijders, T. A., Steglich, C., Schweinberger, M., & Huisman, K. (2008). *Manual for SIENA Version 3.2. ICS*. Groningen, Oxford: University of Groningen, Department of Statistics, University of Oxford.

Snijders, T. A., van de Bunt, G. G., & Steglich, C. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, *32*(1), 44–60. doi:10.1016/j.socnet.2009.02.004.

de Sola Pool, I., & Kochen, M. (1978). Contacts and influence. *Social Networks*, *1*(1), 5–51. doi:10.1016/0378-8733(78)90011-4.

Travers, J., & Milgram, S. (1969). An experimental study of the small world problem. *Sociometry*, *32*(4), 425–443. doi:10.2307/2786545.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. doi:10.1038/30918.

Ziman, J. (1994). *Prometheus bound. Science in dynamic steady state*. Cambridge: Cambridge University Press.