

Scientific collaboration dynamics in a national scientific system

Anuška Ferligoj · Luka Kronegger · Franc Mali ·
Tom Snijders · Patrick Doreian

Version: March 17, 2015

Abstract This paper examines the collaboration structures and dynamics of the co-authorship network of all Slovenian researchers. Its goal is to identify the key factors driving collaboration and the main differences in collaboration behavior across scientific fields and disciplines. Two approaches to modelling network dynamics are combined in this paper: the small-world model and the mechanism of preferential attachment, also known as the process of cumulative advantage. Stochastic-actor-based modelling of co-authorship network dynamics uses data for the complete longitudinal co-authorship networks for the entire Slovenian scientific community from 1996 to 2010. We confirmed the presence of clustering in all fields and disciplines. Preferential attachment is far more complex than a single global mechanism. There were two clear distinctions regarding collaboration within scientific fields and disciplines. One was that some fields had an internal national saturation inhibiting further collaboration. The second concerned the differential impact of collaboration with scientists from abroad on domestic collaboration. In the natural, technical, medical, and biotechnical sciences, this promotes collaboration within the Slovenian scientific community while in the social sciences and humanities this inhibits internal collaboration.

Keywords:

scientific collaboration, co-authorship networks, network dynamics, bibliometry, small world, preferential attachment, stochastic-actor-based model, SIENA, cluster analysis

A. Ferligoj
Faculty of Social Sciences, University of Ljubljana, Slovenia.
E-mail: anuska.ferligoj@fdv.uni-lj.si

L. Kronegger
Faculty of Social Sciences, University of Ljubljana, Slovenia.
E-mail: luka.kronegger@fdv.uni-lj.si

F. Mali
Faculty of Social Sciences, University of Ljubljana, Slovenia.
E-mail: franc.mali@fdv.uni-lj.si

T.A.B. Snijders
University of Groningen, The Netherlands and University of Oxford, UK.
E-mail: t.a.b.snijders@rug.nl

P. Doreian
Faculty of Social Sciences, University of Ljubljana, Slovenia.
Department of Sociology, University of Pittsburgh, Pittsburgh, PA, USA
E-mail: pitpat@pitt.edu

1 Introduction

Scientific collaboration in modern science appears to be one of the key factors for increasing publication productivity and quality. Ziman (1994, p. 218) wrote: "... the traditional parochial individualism of science is rapidly being transformed in what might be described as transnational collectivism". In recent decades, structural changes appearing in science have encouraged scientific collaboration. Specialization at the individual level and the development of very sophisticated and expensive research equipment support collaboration. The development of information and communication technology and increased possibilities for the mobility of researchers also have positive effects on scientific collaboration. Another very important driver of such collaboration is research financing policy. For example, since the EU Framework Programmes (FP) were established in 1984, intra-European collaboration has increased significantly. In addition, the FP Horizon 2020 initiative for research and innovation strongly encourages scientific collaboration.

Scientific collaboration has been studied systematically since the 1960s. Different qualitative and quantitative approaches have been used to study scientific collaboration including interviews, observations, surveys, bibliometric analyses, social network analysis and simulations (e.g., Shrum and Mullins, 1988; Shrum *et al.*, 2007). The most commonly used approach is the bibliometric analysis of scientific co-authorship networks as the data can be easily and accurately extracted from publication databases (Pike, 2010), albeit after considerable cleaning.

Co-authorship networks and citation networks are very useful instruments for studying collaboration in science. Both have positive impacts on scientific productivity. In earlier bibliometric approaches, several studies focused on co-authorship in the social or natural sciences, but very few included comparisons between different scientific disciplines (De Stefano *et al.*, 2011). Scientific disciplines still represent a crucial institutional and organizational framework within which scientific activities take place. They are seen as distinct intellectual and social organizational contexts having their own norms and values forming disciplinary cultures which have evolved over time to create an ever more complicated scientific domain (Kronegger *et al.*, 2014). There are several international, national, and informal classifications of scientific disciplines. Since we analyzed the co-authorship networks of the Slovenian scientific system, we started with the classification into scientific fields and scientific disciplines used by the Slovenian Research Agency (henceforth: ARRS), the main policy authority in Slovenian science. It is presented in Table 1 together with the number of scientific disciplines assigned to each scientific field. The seventh scientific field (Interdisciplinary studies) never gained full recognition as a separate field in the research and development (R&D) policy context in Slovenia because R&D policy remained conservative concerning interdisciplinary-oriented research despite it having increased dramatically around the world. As a result, we ignored it in our analyses of co-authorship network dynamics.

| ID | Scientific field | No. of disciplines |
|----|---------------------------------------|--------------------|
| 1 | Natural sciences and mathematics | 9 |
| 2 | Engineering sciences and technologies | 19 |
| 3 | Medical sciences | 9 |
| 4 | Biotechnical sciences | 6 |
| 5 | Social sciences | 11 |
| 6 | Humanities | 12 |
| 7 | Interdisciplinary studies | 2 |

Table 1 Seven scientific fields in the Slovenian Research Agency's classification system with the number of scientific disciplines

According to analyses of the collaboration styles of researchers belonging to different scientific disciplines and fields, co-authorship represents an important differentiating indicator between them. The differences in the percentages of co-authored publications among the seven scientific fields in Slovenia are shown in Figure 1. There is a large gap between the average levels of co-authorship in the natural, technical, medical and biotechnical sciences, and the average co-authorship levels in the humanities and the social sciences. While six trajectories show a steady rise during 1996-2010, the trajectory for Interdisciplinary studies had considerable

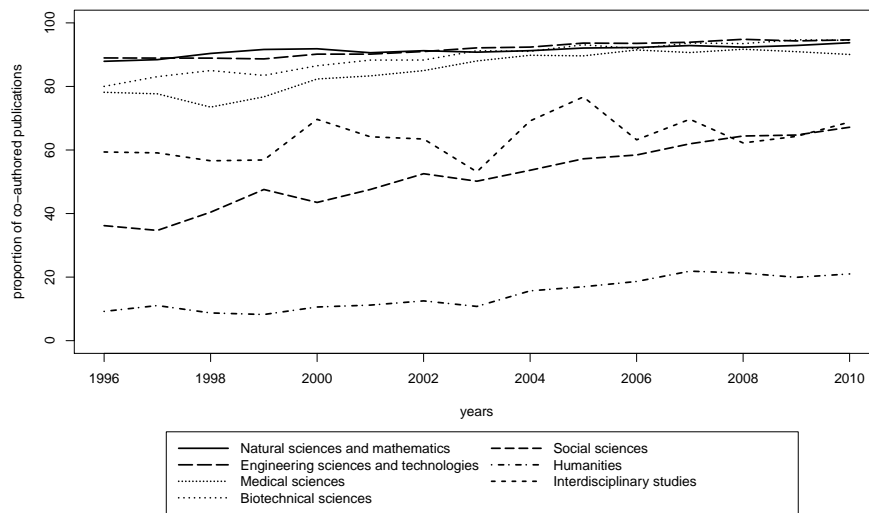


Fig. 1 Percentages of co-authored publications in six scientific fields in Slovenia from 1996 to 2010

instability. The humanities and the social sciences had the smallest shares of co-authored publications throughout this period. However, the trend for the social sciences shows that it is slowly approaching the natural, technical, medical, and biotechnical sciences.

Slovenian researchers have studied scientific collaboration using a qualitative approach, survey analysis and bibliometric analysis of co-authorship networks using longitudinal data on the Slovenian science system in order to explore and explain their dynamics across four scientific disciplines (Ferligoj and Kronegger, 2009; Mali *et al.*, 2010; Kronegger *et al.*, 2011, 2012; Groboljšek *et al.*, 2014; Iglič *et al.*, 2014). The rationale for their selection of four disciplines was: (1) mathematics is an old discipline where research takes place primarily in offices; (2) physics is an old discipline where the research occurs mostly in research groups within laboratories; (3) sociology is an old discipline where research also occurs mostly in offices; and (4) biotechnology is a new laboratory discipline. The bibliometric analysis of co-authorship networks for the period from 1986 to 2005 revealed: (1) a high proportion of single-author publications within sociology and within mathematics with lower but slightly increasing levels of co-authored publications and (2) high and relatively steady levels of collaboration by physicists and biotechnologists within their disciplines and/or with authors from abroad. The high degree of scientific collaboration in laboratory sciences is coupled to a more formal division of labor. Biotechnologists in Slovenia were characterized by collaboration with researchers from other disciplines within the country, although with much greater fluctuations than in the other disciplines (Kronegger *et al.*, 2012).

A web survey of researchers from the four scientific disciplines was designed to understand which kinds of incentives, perceptions, and personal strategies help account for collaboration from the perspectives of individual scientists (Iglič *et al.*, 2014). The results of the analysis showed that the differences between disciplines in the proportion of researchers active work time spent collaborating with others were much smaller when assessed through interviews compared to results from co-authorship data. Researchers in the social sciences do not necessarily collaborate less as the attributions of authorship are different in the social and natural sciences. Disciplines vary according to the nature of collaboration partners. Physicists and mathematicians from the basic sciences have much wider and far-reaching collaboration networks than sociologists or biotechnologists from applied sciences. The extensive collaboration networks of biotechnologists seem to be more limited and focused on local partners.

Disciplines differ also in their overall styles of collaboration, ranging from very informal to formal. For the former, this is due to an established division of labor within an organizational setup with clearly defined roles. Such formal arrangements are far less prevalent in the latter disciplines. Informal collaboration occurs outside an explicit organizational division of labor. The extreme form is when researchers collaborate on their own volition (see Shrum *et al.*, 2007). Iglič *et al.* (2014) also identified the main factors driving collaboration at two levels: research-policy-related external factors and internal factors affecting the motivations of scientists regarding compatibility, cultural proximity, academic excellence, position and status. Scientists are more likely to collaborate when there are more opportunities to obtain research money, when they easily overcome status differences between senior and junior colleagues, and have a high level of agreement about what constitutes good quality research. They also collaborate more when they have a positive experience with earlier collaborations when they experience benefits from collaboration including good research results and faster individual promotion. Also, they pay attention to professional complementarity when choosing research partners.

For an even more detailed analysis of collaboration practices, a qualitative investigation among key representatives of the four scientific disciplines and research policymakers was conducted to gain an insight into their views on the importance of scientific collaboration (Groboljšek *et al.*, 2014). While the policy mechanisms aiming to encourage scientific collaboration are important, the interviewed scientists and policymakers believed that long-term and successful collaborations derive from researcher's efforts and their individual engagements - but only where suitable conditions have been created by policy mechanisms which encourage international mobility, along with interdisciplinary and interinstitutional/intersectoral collaboration.

Kronegger *et al.* (2012) combined two approaches for modelling co-authorship network dynamics. They used the small-world model (Watts and Strogatz, 1998) and the mechanism of preferential attachment, also known as the process of cumulative advantage (Price, 1963, 1965; Garfield and Merton, 1979). One dimension of the small world was measured by its clustering level, and preferential attachment was operationalized through the collaboration of researchers within and across disciplines. They used stochastic-actor-based modelling (SAOM) of co-authorship network dynamics implemented in the SIENA program (Snijders, 2005; Snijders *et al.*, 2010). While the presence of clustering was confirmed in all four scientific disciplines, preferential attachment was more complex than one single global mechanism. The principle of preferential attachment was only partly confirmed in the co-authorship network of physicists: researchers had fewer possibilities to establish new connections with other physicists within the Slovenian scientific 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., in laboratories). The co-authorship network of mathematicians indicated different characteristics: those scientists who collaborated with scientists outside the Slovenian community and had a higher number of articles published in journals with an impact factor had greater opportunities to establish new connections within the Slovenian mathematical community. In the case of the co-authorship networks of sociologists, collaboration with foreign scientists brought negative effects for collaboration within the national research community. At the same time, the number of articles with an impact factor had a positive effect on the formation of new co-authorship relationships within the discipline. The dynamics of biotechnologists co-authorship networks did not follow the principle of preferential attachment.

Here, based on the modelling of co-authorship network dynamics in four scientific disciplines (Kronegger *et al.*, 2012) we specify the SAOM in a more appropriate fashion and use higher quality bibliometric data for the complete longitudinal co-authorship networks for *all* scientific fields and most scientific disciplines for 1996-2010 in the Slovenian system of science.

2 Theoretical arguments and hypotheses

Since the early work of Price (1963, 1965) and Garfield and Merton (1979), sociologists introduced several theories regarding scientific collaboration. Here, we focus on the theory of cumulative advantage in science, termed the Matthew effect (Merton, 1968, 1973; Price, 1976) and the theory of small-world structure (de Sola Pool and Kochen, 1978) and their applications to the modelling of the dynamics of co-authorship networks. In general, we follow the literature in proposing the following mechanisms as influencing scientific collaboration:

(a) network embeddedness: co-authors of co-authors tend to become co-authors; (b) preferential attachment: authors seek out co-authors preferentially who have already many co-authors; (c) institutional embeddedness (belonging to the same research group and to the same scientific discipline, age similarity may also fall under this heading as it means belonging to a common cohort of scholars who interact with each other more than those of different cohorts) and (d) control variables, specifically age, PhD and gender.

Based on the Kronegger *et al.* (2012) study discussed in the previous section which dealt with only four disciplines, we examine whether the earlier findings hold more generally when *whole* fields and *all* disciplines are considered. We test three hypotheses concerning the small world phenomenon, preferential attachment and the impacts of actor agency and institutional contexts.

The small-world model was defined formally by Watts and Strogatz (1998) who introduced an algorithm to construct networks with the following properties: (i) having short paths between any two vertices (and hence, smaller average lengths for the shortest paths) and (ii) incorporating clustering (small dense parts of the network). These properties were later used to identify a small-world structure in co-authorship networks (e.g., Newman, 2000, 2001; Moody, 2004; Perc, 2010; Çavuşoğlu and Türker, 2013, 2014). In this respect, Perc (2010) examined the entire Slovenian system of science for 1965–2010. He focused on the largest component, the clustering coefficient, and the mean distance between authors, all of which are consistent with the small-world model. He showed also the network in Slovenia is growing exponentially. Here, the hypothesis dealing with the clustering level, the second property of the small-world model, is:

H_1 : The co-authorship networks in the Slovenian scientific community have a high clustering level driven by transitive closure processes where co-authors of co-authors become, or remain, co-authors.

The idea of cumulative advantage implies that excellent scientists are rewarded far more than others in their field. Said *et al.* (2008) noted one factor affecting co-authorship ties is the mentor-student relationship: young researchers are more likely to form new co-authorship ties with older, established researchers, usually their mentors. The formal modelling of cumulative advantage in terms of preferential attachment as the driving mechanism of co-authorship was examined by Barabási and Albert (1999) who studied a common property of many large networks whose vertex degrees followed a scale-free power-law distribution. This feature was found to be a consequence of two generic mechanisms: (i) networks expand continuously with 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 properties and reproduced the observed stationary scale-free distributions. The model was widely accepted and also criticized (e.g. Wagner and Leydesdorff, 2005; Abbasi *et al.*, 2012). Li *et al.* (2006) provided a general assessment of scale-free distributions including instances where such distributions fail to capture network features. They also lay the foundations for fruitful applications of scale-free networks. The implications of scale-free distributions were used to delineate the structure of scientific co-authorship networks (e.g. Barabási, 2002; Moody, 2004; Perc, 2010; Kronegger *et al.*, 2011). We examine these arguments further by asking if authors who are already well connected, as evidenced by their current number of co-authorships, will attract even more co-authorships as time goes by.

As noted earlier, the concept of preferential attachment reduces the generation of co-authorship to a single mechanism. However, the phenomenon of collaboration is far more complex. This led Kronegger *et al.* (2012) to test for the presence of preferential attachment in Slovenian co-authorship networks over time in four scientific disciplines. Using SAOM, they operationalized preferential attachment by separating the collaboration of researchers within the scientific discipline from their collaboration with scientists from abroad. They showed that some features of the preferential attachment principle were confirmed but in *different ways* in the three considered scientific disciplines, and *not at all* in biotechnology. Here we test the following hypothesis dealing with the preferential attachment mechanisms using higher quality data:

H_2 : New co-authorship collaborations of Slovenian researchers are more likely for authors who have more current co-authorships and for excellent researchers. For co-authorships, this holds both for collaboration within Slovenia and with researchers abroad.

The hypothesis that individual and organizational contexts drive the formation of scientific co-authorship networks was confirmed by Kronegger *et al.* (2012). They showed that the four disciplines were affected in different ways by the organization of local institutions and disciplinary publishing cultures. Here, based on

their analyses we test the same hypothesis for *all* scientific fields and most scientific disciplines in the Slovenian scientific system.

H_3 : Individual and organizational contexts in Slovenia drive the formation of scientific co-authorship networks.

3 Model specification

The three hypotheses were tested using an actor-oriented model (Snijders, 2001, 2005; Snijders *et al.*, 2007, 2010) used for longitudinal network data with a model defined as continuous-time Markov process. Since our data are non-directed networks, a modification to the models of Snijders (2001, 2005) is required. To obtain a non-directed network, the assumption is made that at random moments, a randomly chosen actor ('ego') chooses another actor ('alter') to propose a new tie or to drop an existing tie; if a new tie is proposed, alter can decide to accept or reject the proposal (see Snijders, 2008). The choice by ego of alter is a multinomial choice, and the acceptance decision by alter is a binary choice. The probability models for these choices are based on a linear predictor similar to generalized linear models. The coefficients parameters given in Table 4 are the estimated parameters in these linear predictors.

Given the first hypothesis, we included in the model a clustering component to capture the idea of small dense parts being present in the network. As clustering can be viewed as a consequence of transitive closure, we added to the model the effect of transitivity in triplets. Co-authors of co-authors will have a larger probability to become direct co-authors; and if they are already direct co-authors, they will have a larger probability of remaining compared to pairs of authors who are not co-authors of co-authors. Also, as co-authorship can also be driven by departmental and institutional affiliation, we operationalized this by working in the same organizational research group and by working in the same scientific discipline when analyzing fields.

Regarding H_2 , there are several options for capturing preferential attachment. In the first place, it is tested whether the current degree (number of co-authorships) has a positive effect on the number of new co-authorships. Since the collaboration network is symmetric, there is no distinction between the actors at both sides of the tie. Therefore, individual variables are included without an ego-alter differentiation. As the degree captures only collaborations inside national networks (scientific fields or scientific disciplines on national level), we also included collaboration outside the national collaboration network in the model. As this variable was highly skewed we used its logarithm. Scientific excellence was measured by a dichotomous variable where 1 represented that the researcher has at least one publication published in the top international scientific journals according to the journals classification of the Slovenian Research Agency (see <https://www.arrs.gov.si/en/akti/prav-sof-ocen-sprem-razisk-dej-sept-11.asp>). The organizational context for H_3 was operationalized by the variables 'working in the same research group' and 'working in the same scientific discipline'. Individual context was considered by scientific excellence. To study how young researchers form new co-authorship ties with older established researchers we included into the model 'scientific age' (defined as the year of an author's first publication) and 'age similarity'. As controlling variables, we included gender and having a PhD.

4 Data

Our analyses were performed on the bibliographic data of all Slovenian researchers, scientific fields and disciplines to which they belonged in the period 1996–2010. In the first five years after Slovenias independence, governmental institutions made many efforts to establish the new country and new legislation, including the sphere of science. From 1996 on, some stability of the scientific system can be observed. This can be seen in Figure 1 showing the percentages of co-authored publications in the six scientific fields. For a subsequent more fine-grained temporal image, we analyzed the co-authorship networks starting five years after the independence of Slovenia. The data were organized in three 5-year intervals:

- Period 1, 1996–2000: a period of harmonization with the European Union (EU) and the OECD standards;

- Period 2, 2001–2005: in 2004, Slovenia became a member of the EU. The Slovenian Research Agency was established in the same year followed by many positive effects on R&D evaluation procedures due to its policies; and
- Period 3, 2006–2010: a more stable period.

The dataset was obtained from the Current Research Information System (SICRIS) which includes information on all current and former researchers registered with the Slovenian Research Agency and with the co-operative On-Line Bibliographic System & Services (COBISS) which is an officially maintained database of all publications available in Slovenian libraries. From this system, we collected complete scientific bibliographies of all Slovenian researchers who had ever been given a research identification number (ARRS ID) by the Slovenian Research Agency. This requirement for selecting researchers in a specific scientific field or discipline differs from the study from Kronegger *et al.* (2012) who considered only researchers in four scientific disciplines that were included in 2008 in the SICRIS database.¹ As a result, we obtained much larger networks, e.g. for Physics during 1996–2000 183 researchers were considered in the earlier paper whereas, here, we have 288 physicists. For mathematics there were 96 researchers in the earlier study and 146 in the present analysis. For Biotechnology there were 50 researchers and now 106 and, for Sociology, we have 129 compared to 88 in the earlier study.

The total number of researchers with an ARRS ID who published in the time period 1996–2010 was 15,424. These researchers collaborated with another 48,191 authors not registered with ARRS. Together, they published 170,118 publications that are, according to the evaluation criteria of ARRS, treated as scientific outputs. The data about discipline memberships were provided by the researchers themselves when they applied for an identification number.

As noted above, the Slovenia's national scientific system is organised into 72 scientific disciplines classified into 7 research fields. There were 14 disciplines (with their identification numbers in parentheses) excluded for the following reasons:

- Technology driven physics (30), Communications technology (31), Landscape design (45) and Ethnic studies (57) were excluded because of having too small numbers of researchers.
- Anthropology (62), Culturology (65), Literary (66), Musicology (67), Philosophy (69), and Theology (70) were excluded also due to having few researchers (all less than 30) and few co-publications (average degrees less than 1), and high turnover in co-authorship (authors in one year vanish and are replaced by new authors). This was measured by Jaccard coefficients between consecutive periods.
- Law (51) was excluded due to a deviating data structure: each wave had a few papers having a very high number of authors in contrast to the usual number of authors for this discipline.
- Historiography (60) was excluded due to high proportion of missing values in variables for actor properties.
- The NCKS Research programme (72), in addition to Interdisciplinary research (73), was excluded as it lacks an established field structure (Interdisciplinary studies).

Our analyses were performed for six scientific fields and 58 research disciplines out of 72. Table 2 has summary results for all scientific fields and Table 3 provides the corresponding results for all disciplines. For each time period we provide the numbers of researchers, average degrees and the number of researchers having at least one co-author in their publications during the period 1996–2010 (i.e. were connected) within their scientific field or scientific discipline. In all scientific fields and in most scientific disciplines the number of researchers grew through the three time periods considered. Only 13 disciplines had no strict increase in the number of researchers. Larger fluctuations in the number of researchers in the three time periods were evident in the following disciplines: Electric devices; Process engineering; Textile and leather; Metrology; Mining and geotechnology and Veterinarian medicine.

The last column in the last panel in both Tables 2 and 3 presents the share of researchers not having any co-authored publication during the observed time period within each scientific field or discipline. For fitting models to these data, there are two options for treating these isolates in the co-authorship networks: a new parameter called 'isolates' could be included in the model or the isolates could be excluded. As both approaches

¹ Therefore, some older, mostly prominent researchers were missing along with some other researchers.

| No. | Name | Field | Period | | | Average degree | | | Researchers | | |
|-----|---------------------------------------|-------|--------|------|------|----------------|------|------|-------------|-----------|------------|
| | | | 1 | 2 | 3 | 1 | 2 | 3 | all | connected | % excluded |
| 1 | Natural sciences and mathematics | | 1538 | 1795 | 2089 | 2.80 | 3.80 | 5.13 | 2585 | 2294 | 11 |
| 2 | Engineering sciences and technologies | | 2355 | 2649 | 2994 | 2.47 | 3.25 | 4.44 | 4040 | 3762 | 7 |
| 3 | Medical sciences | | 1470 | 1636 | 1720 | 4.53 | 5.77 | 6.62 | 2144 | 1978 | 8 |
| 4 | Biotechnical sciences | | 769 | 797 | 919 | 3.03 | 4.44 | 5.99 | 1192 | 1108 | 7 |
| 5 | Social sciences | | 1309 | 1648 | 1830 | 1.76 | 2.34 | 3.23 | 2193 | 1718 | 22 |
| 6 | Humanities | | 996 | 1226 | 1350 | 0.39 | 0.60 | 1.30 | 1556 | 736 | 53 |

Table 2 Basic characteristics in networks of research fields

give the same results, the isolates were excluded from all further analyses.² There is considerable variation in the percentages of excluded researchers across scientific fields where the excluded researchers were single authors or researchers who wrote publications mostly with researchers from abroad.³ As expected, there were small percentages (between 7 and 11%) of such researchers in the natural, technical, medical, and biotechnical sciences. The social sciences had a much larger percentage (22%) with the largest percentage (53%) in the humanities where there were more solo-authored publications. This tendency is also seen in the middle panel of Tables 2 and 3 (average degrees).

A similar trend is evident in the scientific disciplines (see Table 3) where the variation between the percentages of excluded researchers across scientific disciplines inside each of the scientific fields is larger. In the natural sciences, there are four disciplines (out of nine) with percentage exceeding 30% (Mathematics, Geology, Computer intensive methods and applications, and Control and care of the environment).⁴ There is less variation in the engineering sciences and technology and in the medical sciences. The only outlier is Technology driven physics with an extremely small number of researchers. As noted above, there are higher percentages of excluded researchers in the disciplines from the social sciences and the humanities.

The scientific disciplines listed at the bottom of Table 3 and marked by EXC were excluded from all further analyses. All co-authorship ties were binarized: if two researchers had at least one joint publication a value of 1 was assigned, otherwise a value of 0 was used.

Two control variables, gender (male=1) and having a PhD (having PhD=1), were used.

5 Results

The preferential mechanism described by Barabási and Albert (1999) focuses on the formation of co-authorship ties as new nodes enter the network. It appears that, in their analyses, such ties remain in the network, which may be problematic in the case of co-authorship networks. Yet, scientists collaborating at one point in time can maintain or dissolve their co-authorship tie at a later time. Consequently, we considered the possibility that ties can be created, maintained or deleted since this is a feature that characterizes co-authorship networks. This was a major reason for our use of Stochastic Actor-Oriented Model (SAOM; Snijders 2001, 2008), a micro-level model assuming that ties can be created and deleted in a network with an actor set that may be constant or changing. The probabilities of tie creation and deletion depend on so-called effects (explanatory variables that may depend on the network or be exogenously given), with their associated coefficients as parameters. The set of effects chosen as the model specification followed from the hypotheses as elaborated in the preceding section.⁵ The clustering level was operationalized by the parameter for transitive triads to capture the tendency of actors to form ties by closing the triangles, and by the parameters ‘working in the same research organizational group’ and ‘working in the same scientific discipline’. The last two are a better operationalization of the institutional feature of the clustering level than the one used in the study by Kronegger *et al.* (2012). Preferential attachment was operationalized by five effects: degree of alter inside the co-authorship network, degree

² The resulting networks were smaller and the SIENA analyses were more efficient.

³ There is a small percentage of co-authored publications coming from different scientific fields (see Kronegger *et al.*, 2014) or outside the ARRS system inside Slovenia.

⁴ In other classification systems, some of these disciplines are not classified in the natural sciences.

⁵ A similar formalization of the stochastic-actor-based model for modelling co-authorship networks was used by Kronegger *et al.* (2012).

| Disc. No. | Status | Field number and discipline name | Period | | | Average degree | | | Researchers | | |
|-----------|--------|---|--------|-----|-----|----------------|------|------|-------------|-----------|------------|
| | | | 1 | 2 | 3 | 1 | 2 | 3 | all | connected | % excluded |
| 1 | | 1 Mathematics | 146 | 186 | 241 | 0.56 | 1.08 | 1.94 | 269 | 178 | 34 |
| 2 | | 1 Physics | 288 | 341 | 421 | 2.92 | 3.90 | 5.63 | 478 | 417 | 13 |
| 3 | | 1 Biology | 251 | 313 | 395 | 1.02 | 1.67 | 2.68 | 450 | 328 | 27 |
| 4 | | 1 Chemistry | 406 | 427 | 477 | 2.88 | 3.24 | 3.48 | 624 | 553 | 11 |
| 5 | | 1 Biochemistry and molecular biology | 158 | 216 | 265 | 1.89 | 2.43 | 2.64 | 318 | 221 | 31 |
| 6 | | 1 Geology | 196 | 236 | 270 | 1.69 | 2.16 | 2.22 | 321 | 193 | 40 |
| 7 | | 1 Computer intensive methods and applications | 29 | 50 | 60 | 0.12 | 0.66 | 1.10 | 67 | 38 | 43 |
| 8 | | 1 Control and care of the environment | 50 | 77 | 90 | 0.25 | 0.59 | 0.89 | 122 | 62 | 49 |
| 9 | | 1 Pharmacy | 187 | 206 | 228 | 1.93 | 2.56 | 3.85 | 326 | 275 | 16 |
| 10 | | 2 Civil engineering | 216 | 215 | 245 | 1.52 | 1.98 | 2.75 | 334 | 295 | 12 |
| 11 | | 2 Chemical engineering | 135 | 143 | 160 | 1.40 | 1.80 | 2.19 | 225 | 184 | 18 |
| 12 | | 2 Energy engineering | 183 | 218 | 241 | 1.68 | 2.15 | 2.94 | 310 | 263 | 15 |
| 13 | | 2 Materials science and technology | 213 | 251 | 279 | 2.43 | 3.48 | 4.45 | 368 | 313 | 15 |
| 14 | | 2 Mechanics | 51 | 58 | 57 | 0.54 | 0.80 | 1.04 | 85 | 48 | 44 |
| 15 | | 2 Systems and cybernetics | 212 | 222 | 257 | 2.48 | 2.83 | 3.26 | 331 | 275 | 17 |
| 16 | | 2 Computer science and informatics | 366 | 411 | 485 | 1.54 | 1.95 | 2.72 | 645 | 509 | 21 |
| 17 | | 2 Telecommunications | 127 | 133 | 159 | 1.16 | 1.98 | 3.12 | 232 | 182 | 22 |
| 18 | | 2 Electronic components and technologies | 99 | 102 | 111 | 1.72 | 1.47 | 2.37 | 163 | 128 | 21 |
| 19 | | 2 Manufacturing technologies and systems | 186 | 193 | 232 | 1.22 | 1.68 | 2.85 | 322 | 246 | 24 |
| 20 | | 2 Mechanical design | 109 | 154 | 172 | 1.05 | 2.02 | 2.26 | 232 | 194 | 16 |
| 21 | | 2 Electric devices | 60 | 79 | 70 | 1.69 | 2.64 | 3.61 | 109 | 82 | 25 |
| 22 | | 2 Process engineering | 80 | 75 | 81 | 1.98 | 1.72 | 2.07 | 109 | 92 | 16 |
| 23 | | 2 Textile and leather | 86 | 109 | 99 | 1.33 | 2.63 | 3.07 | 149 | 133 | 11 |
| 24 | | 2 Metrology | 66 | 85 | 76 | 1.20 | 1.24 | 1.33 | 113 | 84 | 26 |
| 25 | | 2 Mining and geotechnology | 34 | 30 | 26 | 1.02 | 1.38 | 1.51 | 45 | 36 | 20 |
| 26 | | 2 Geodesy | 33 | 47 | 56 | 1.15 | 1.87 | 4.92 | 61 | 55 | 10 |
| 28 | | 2 Traffic systems | 49 | 60 | 73 | 0.91 | 1.30 | 1.95 | 86 | 72 | 16 |
| 29 | | 2 Hydrology | 38 | 45 | 59 | 1.41 | 2.84 | 4.31 | 64 | 53 | 17 |
| 32 | | 3 Microbiology and immunology | 199 | 231 | 234 | 4.49 | 4.83 | 5.1 | 288 | 263 | 9 |
| 33 | | 3 Stomatology | 52 | 52 | 55 | 1.49 | 2.00 | 2.65 | 74 | 62 | 16 |
| 34 | | 3 Neurobiology | 276 | 308 | 306 | 1.54 | 2.18 | 2.84 | 386 | 331 | 14 |
| 35 | | 3 Oncology | 216 | 238 | 245 | 4.32 | 5.24 | 5.02 | 299 | 262 | 12 |
| 36 | | 3 Human reproduction | 157 | 171 | 149 | 3.61 | 5.25 | 4.82 | 204 | 182 | 11 |
| 37 | | 3 Cardiovascular system | 216 | 222 | 239 | 3.24 | 3.45 | 3.79 | 291 | 256 | 12 |
| 38 | | 3 Metabolic and hormonal disorders | 71 | 82 | 89 | 0.60 | 0.90 | 1.50 | 107 | 74 | 31 |
| 39 | | 3 Public health (occupational safety) | 180 | 193 | 234 | 0.82 | 0.94 | 1.70 | 301 | 215 | 29 |
| 40 | | 3 Psychiatry | 28 | 44 | 46 | 1.25 | 1.85 | 3.66 | 53 | 47 | 11 |
| 41 | | 4 Forestry, wood and paper technology | 132 | 122 | 136 | 1.99 | 3.03 | 3.65 | 181 | 162 | 10 |
| 42 | | 4 Animal production | 119 | 127 | 130 | 2.88 | 5.10 | 4.45 | 176 | 164 | 7 |
| 43 | | 4 Plant production | 224 | 235 | 280 | 2.38 | 3.66 | 4.76 | 355 | 320 | 10 |
| 44 | | 4 Veterinarian medicine | 157 | 141 | 143 | 4.68 | 5.32 | 5.65 | 203 | 191 | 6 |
| 46 | | 4 Biotechnology | 106 | 130 | 173 | 1.37 | 2.05 | 4.62 | 211 | 185 | 12 |
| 47 | | 5 Educational studies | 270 | 334 | 343 | 1.30 | 1.54 | 2.08 | 427 | 295 | 31 |
| 48 | | 5 Economics | 308 | 393 | 420 | 1.11 | 2.12 | 2.79 | 521 | 405 | 22 |
| 49 | | 5 Sociology | 129 | 163 | 179 | 2.48 | 2.79 | 2.66 | 210 | 149 | 29 |
| 50 | | 5 Administrative and organisational sciences | 84 | 109 | 140 | 0.34 | 0.43 | 0.93 | 163 | 92 | 44 |
| 52 | | 5 Political science | 92 | 116 | 132 | 0.61 | 1.93 | 1.96 | 150 | 105 | 30 |
| 53 | | 5 Criminology and social work | 46 | 58 | 67 | 0.92 | 1.30 | 3.76 | 74 | 65 | 12 |
| 54 | | 5 Urbanism | 40 | 42 | 51 | 0.34 | 0.55 | 0.86 | 65 | 42 | 35 |
| 55 | | 5 Psychology | 80 | 97 | 123 | 0.51 | 1.10 | 2.16 | 134 | 88 | 34 |
| 56 | | 5 Sport | 17 | 22 | 34 | 0.83 | 0.83 | 3.00 | 36 | 30 | 17 |
| 58 | | 5 Architecture and Design | 51 | 56 | 63 | 0.24 | 0.31 | 1.34 | 85 | 51 | 40 |
| 59 | | 5 Information science and librarianship | 24 | 32 | 34 | 0.08 | 0.42 | 1.00 | 48 | 26 | 46 |
| 61 | | 6 Archaeology | 51 | 69 | 70 | 0.68 | 1.22 | 1.70 | 79 | 66 | 16 |
| 63 | | 6 Ethnology | 54 | 66 | 61 | 0.05 | 0.67 | 0.79 | 78 | 27 | 65 |
| 64 | | 6 Linguistics | 186 | 238 | 276 | 0.21 | 0.43 | 0.83 | 308 | 141 | 54 |
| 68 | | 6 Art history | 46 | 61 | 72 | 0.05 | 0.15 | 1.65 | 79 | 34 | 57 |
| 71 | | 6 Geography | 78 | 94 | 106 | 1.21 | 1.08 | 2.60 | 124 | 96 | 23 |
| 30 | EXC | 2 Technology driven physics | 1 | 1 | 5 | 0.00 | 0.00 | 0.00 | 5 | 0 | 100 |
| 31 | EXC | 2 Communications technology | 6 | 10 | 23 | 0.00 | 0.17 | 2.43 | 23 | 17 | 26 |
| 45 | EXC | 4 Landscape design | 25 | 24 | 26 | 1.66 | 0.46 | 0.63 | 35 | 21 | 40 |
| 51 | EXC | 5 Law | 145 | 182 | 181 | 2.45 | 2.40 | 2.19 | 211 | 139 | 34 |
| 57 | EXC | 5 Ethnic studies | 2 | 4 | 7 | 0.00 | 0.00 | 0.29 | 7 | 2 | 71 |
| 60 | EXC | 6 Historiography | 173 | 200 | 222 | 0.20 | 0.64 | 2.18 | 266 | 103 | 61 |
| 62 | EXC | 6 Anthropology | 67 | 76 | 79 | 0.12 | 0.14 | 0.14 | 97 | 20 | 79 |
| 65 | EXC | 6 Culturology | 54 | 68 | 77 | 0.09 | 0.07 | 0.21 | 85 | 21 | 75 |
| 66 | EXC | 6 Literary sciences | 102 | 119 | 120 | 0.09 | 0.10 | 0.09 | 139 | 29 | 79 |
| 67 | EXC | 6 Musicology | 24 | 27 | 34 | 0.11 | 0.05 | 0.00 | 38 | 6 | 84 |
| 69 | EXC | 6 Philosophy | 76 | 87 | 85 | 0.08 | 0.12 | 0.14 | 103 | 25 | 76 |
| 70 | EXC | 6 Theology | 21 | 31 | 36 | 0.00 | 0.34 | 0.39 | 41 | 12 | 71 |
| 72 | EXC | 7 NCKS Research programme | 7 | 4 | 4 | 0.00 | 0.00 | 0.00 | 8 | 0 | 100 |
| 73 | EXC | 7 Interdisciplinary research | 13 | 18 | 36 | 0.11 | 0.16 | 0.47 | 38 | 10 | 74 |

Table 3 Basic characteristics in networks of research disciplines

of alter with respect to collaboration with researchers outside the co-authorship network, scientific excellence, scientific age, and age similarity. Gender (male=1) and having a PhD (having PhD=1) were also included in the model as controls. The network was defined in three consecutive observations corresponding to the periods mentioned in Section 4, and a tie was defined if two researchers appeared together as authors in at least one publication.

5.1 Scientific fields

Table 4 presents the estimated parameters for the six scientific fields labeled as Nat (Natural sciences and mathematics), Eng (Engineering sciences and technologies), Med (Medical sciences), Bio (Biotechnical sciences), Soc (Social sciences) and Hum (Humanities). Also reported are the standard errors of these estimates. The shaded estimates are the only ones not statistically significant. A similar table for all of the considered scientific disciplines is in the Appendix (Table 6). The first three parameters in the tables are standard as they are technical requirements of the stochastic-actor-based model: the rate parameter for the first transition; the rate parameter for the second transition, and the density parameter. The two rate parameters estimate the average frequency of the number of proposals for collaboration, which then may be accepted or rejected by the proposed partner within the co-authorship networks. In all scientific fields, the average number of proposals for collaboration for the second transition is higher than for the first transition. This is not the case in all scientific disciplines considered (see Table 6): 12 out of 49 considered disciplines had declines in their averages in the second transition compared to the first one.

| | 1 Nat | 2 Eng | 3 Med | 4 Bio | 5 Soc | 6 Hum |
|------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| rate 1 | 27.255 (0.986) | 23.832 (0.731) | 50.476 (1.574) | 33.287 (1.539) | 41.012 (5.188) | 25.861 (4.052) |
| rate 2 | 32.837 (1.539) | 33.080 (0.895) | 60.107 (1.510) | 35.844 (1.090) | 44.005 (1.758) | 28.612 (9.312) |
| degree (density) | -2.509 (0.031) | -2.649 (0.027) | -2.137 (0.016) | -1.919 (0.034) | -2.463 (0.029) | -2.534 (0.303) |
| transitive triads | 0.377 (0.008) | 0.544 (0.010) | 0.309 (0.008) | 0.314 (0.011) | 0.383 (0.012) | 1.097 (0.108) |
| same research group | 1.131 (0.048) | 1.483 (0.042) | 0.979 (0.025) | 0.668 (0.042) | 0.980 (0.034) | 0.922 (0.181) |
| same discipline | 0.698 (0.016) | 0.848 (0.019) | 0.413 (0.015) | 0.709 (0.023) | 0.723 (0.018) | 0.900 (0.066) |
| degree of alter | -0.014 (0.002) | -0.040 (0.003) | -0.012 (0.002) | -0.022 (0.004) | -0.033 (0.003) | -0.085 (0.087) |
| degree out | 0.150 (0.012) | 0.150 (0.013) | 0.140 (0.010) | 0.110 (0.017) | -0.049 (0.019) | -0.121 (0.082) |
| excellence | -0.168 (0.032) | -0.016 (0.026) | -0.019 (0.023) | -0.059 (0.033) | 0.459 (0.028) | 0.447 (0.195) |
| first publication year | 0.010 (0.001) | 0.007 (0.001) | 0.018 (0.001) | 0.015 (0.001) | 0.006 (0.002) | 0.004 (0.005) |
| first pub. similarity | 0.091 (0.056) | -0.081 (0.087) | 0.037 (0.059) | 0.121 (0.090) | 0.058 (0.071) | -0.011 (0.220) |
| PhD (yes) | 0.780 (0.043) | 0.769 (0.043) | 0.576 (0.019) | 0.463 (0.034) | 0.638 (0.051) | 0.179 (0.133) |
| gender (male) | 0.016 (0.027) | 0.146 (0.029) | 0.165 (0.020) | 0.130 (0.028) | -0.186 (0.028) | 0.149 (0.094) |

Table 4 Estimated parameters for the six scientific fields (shaded estimates are not statistically significant; there are standard errors in parentheses)

Using SAOM includes also an estimate of the cost of adding one more tie to the personal network of each researcher is obtained, an important characteristic seldom considered by scholars studying preferential attachment. This value is given by the third basic parameter (degree).⁶ Estimates of the parameters for the degree effect are *negative* for *all* six scientific fields and *all* scientific disciplines. This makes sense because tie formation incurs costs in terms of time, effort, and resources. Researchers can co-author with only a limited number of different authors as each new tie represents an additional time and cost burden.

The next three parameters in Table 4 concern the clustering level as a dimension of the small-world process. Among these, the fourth parameter of the model, for the transitive triads effect, is positive and significant showing that scientists tend to form new co-authorship ties with the co-authors of their co-authors inside the scientific field. It is positive and significant for *all* fields and *all* disciplines. The fifth and the sixth parameter

⁶ Abbasi *et al.* (2012) have an interesting result showing that betweenness centrality has predictive value regarding the formation of coauthorship ties in an evolutionary perspective. Preferential attachment, in its simple form, uses degree centrality to capture an aspect of researcher motivation for seeking new collaborative ties. For individuals having no interest in network analysis, it is highly unlikely that they are aware of their betweenness values in a network. We did not include betweenness as a predictor for this reason. Even so, this idea is worth considering in future research.

that show the impact of belonging to the same research group and to the same discipline as a tendency to form new co-authorship ties are also positive and significant in five scientific fields. This holds for *all but two* scientific disciplines.⁷ These estimates provide irrefutable confirmation of a high level of clustering within co-authorship networks in the Slovenian scientific community.⁸ The first hypothesis is confirmed emphatically.

The next three parameters in Table 4 concern preferential attachment in the six scientific fields in Slovenia. As discussed in the previous section, this deals with a preference to create new ties with prominent researchers who already have a high number of co-authors. In our model we included the following indicators to measure individual preferential attachment: (i) the alters degree within the co-authorship network, indicating the number of co-authors within the national borders of the field, implemented as an endogenous degree effect; (ii) the logarithm of the alter's degree coming from researchers outside the field and (iii) the alter's publication excellence (operationalized as having at least one paper in the top 25% of journals in a field). The parameter for alters number of co-authors within the scientific field network is negative and statistically significant in all scientific fields (it is negative but not significant in the humanities). This indicates that *researchers do not tend to form new ties with those researchers who collaborate more within the national field*. This is a partial contradiction of the second hypothesis.

There is much greater diversity in the alters number of co-authors outside the scientific field. The parameters showing significant positive values are for Natural sciences and mathematics, Engineering sciences and technologies, Medical sciences, and Biotechnical sciences. A positive estimated parameter means that researchers are more likely to create new ties with those researchers within the field who collaborate with many authors from other fields or mostly with others from abroad. This is partial confirmation for the second hypothesis. However, for Social science and Humanities, the sign of this parameter is negative (significant for the social sciences and not significant for the humanities): among these researchers, collaboration with other researchers (outside the field or outside Slovenia) has a negative effect on tie formation with scientists working in the social sciences and in the humanities. This is another partial contradiction of H_2 . Yet publication excellence has a positive and significant effect on new tie formation within the Social sciences and the Humanities: researchers tend to create new co-authorship ties more often with those researchers who publish in the highest ranked scientific journals. We note that researchers from these two fields publish less often in the best scientific journals suggesting excellence is more valued for creating co-authorship ties. The only negative and significant parameter for publishing excellence is in the Natural sciences and mathematics. This parameter is negative but not significant for other three fields.

In summary, regarding preferential attachment, the social sciences and the humanities contradict H_2 while it is supported in the remaining scientific fields when considering the parameter 'publishing out of the field'.⁹ The parameter for 'degree of alter' is significant and *negative for all fields*, flatly contradicting H_2 . Regarding publication excellence, the results are mixed. It has no effect on co-authorship in the technical, medical and biological sciences. It has a negative effect in the natural sciences. For these four fields, this aspect of the preferential attachment hypothesis is not confirmed.¹⁰ However, the effect of excellence on tie formation is positive in the social sciences and the humanities.

We also tested whether young researchers form new co-authorship ties with older, established researchers (usually their mentors). The coefficient for (scientific) age, operationalized as the year of authors' first publication, is positive and significant for all fields except the humanities. The most salient effects are in the natural sciences and mathematics. The age similarity is not significant in all scientific fields. This result does *not* follow the standard hypothesis claiming young researchers form new co-authorship ties with scientifically excellent older scientists.

⁷ It is negative (but not significant) only in Psychiatry and Textile and leather.

⁸ Similar results for the clustering level in four selected scientific disciplines in Slovenia were also obtained by Kronegger *et al.* (2012).

⁹ Primarily, this is collaboration with researchers from abroad since there are very little collaboration between different fields in Slovenia.

¹⁰ These results are not completely congruent with the results obtained by Kronegger *et al.* (2012) for four scientific fields. The difference can be attributed to having, in this study, larger scientific groups with better data and a more elaborated model specification.

Next, we examine the effects of the controlling variables. As expected, having a PhD has a positive effect on tie formation in all scientific fields. Researchers are more likely to establish ties with male colleagues within the technical, medical and biotechnical sciences. The only negative effect is in the social sciences: researchers are more likely to create new ties with female researchers within the field. This effect is not significant in the natural sciences and the humanities. No doubt this reflects demographic differences: the proportion of female researchers is the highest in the social sciences.

A general summary of the results of fitting the stochastic-actor-oriented model is straightforward. In all six scientific fields, Slovenian researchers form new co-authorship ties in ways consistent with clustering inside the co-authorship networks: co-authors of co-authors will tend to become co-authors. The preferential attachment mechanism is more complex than the advocates of a single global autonomous mechanism claim. First, the distance between the researchers who collaborate matters for tie formation. Alters high degree of co-authorship inside the field has a negative effect on new tie formation in all scientific fields, but high alters degree of collaboration outside the field reveals a gap between the social sciences and humanities and the other four fields. Alters higher degree of outside collaboration has a negative effect on new tie formation and publication excellence has a positive effect in the social sciences and humanities but the opposite effect exists in the other four fields.

5.2 Scientific disciplines

The foregoing results hold for the scientific fields. The next obvious question is whether these results hold for the scientific disciplines within these fields. For this purpose, we estimated the SAOM models for the disciplines. These results are reported fully in the Appendix. In doing this we took another look at the classification of disciplines. Often, some scientific disciplines were assigned to the scientific fields for historical reasons and are classified differently in most other international systems. For instance, Geography is classified in the Humanities in the Slovenian classification system (see Kronegger *et al.*, 2014). Rather than using this system, we opted to cluster the 59 disciplines listed above according to the obtained estimated parameters of the specified stochastic-actor-based model using the clustering level (transitive triads and belonging to the same research group) and the preferential attachment (degree of alter inside the discipline, degree of alter outside the discipline and alter's publication excellence).

Doing this was not straightforward because the estimated parameters are not directly comparable across disciplines due to variations in the size of the disciplines. While the starting point is the set of these estimated parameters, we transformed them to measure the importance of the estimated parameters using the proposed method of Indlekofer and Brandes (2013). Denoting the importance of estimated parameter, j , by I_j , these measures are constrained as $0 \leq I_j \leq 1$. These values ignore the sign of the estimated parameters for disciplines. For disciplines having negative estimated parameters, the sign of the importance measure was multiplied by -1. These measures were standardized before obtaining the Euclidean distances for each pair of disciplines. The clustering used Wards hierarchical clustering procedure (Ward, 1963).¹¹ The obtained dendrogram is shown in Figure 2.

Five clusters were identified. The top cluster is composed of four disciplines from the technical and natural sciences. Computer intensive methods comes from the first field identified in Table 1 while the remaining three come from the second field. We label this cluster ENG,NAT in Table 5. The second cluster contains disciplines from the technical and medical fields. All but one¹² of the disciplines clustered here come from the second and third scientific fields. In Table 5, this is labeled as ENG,MED. The third cluster is composed solely with disciplines from the social science and humanities. This is labeled as HUM,SOC. All of these three clusters are quite homogenous. The fourth cluster is less so. In the main, it is composed of disciplines that can be viewed

¹¹ Alternative transformations were examined also. In one of them, we assigned a value of 1 to the parameters that are positive and statistically significant, a value of -1 to the negative and statistically significant ones and 0 to the non-significant ones before applying this clustering method. Despite the problems of relying on 'significance' which can be affected by sample sizes, the obtained partition was comparable.

¹² Criminology and social work is the exception.

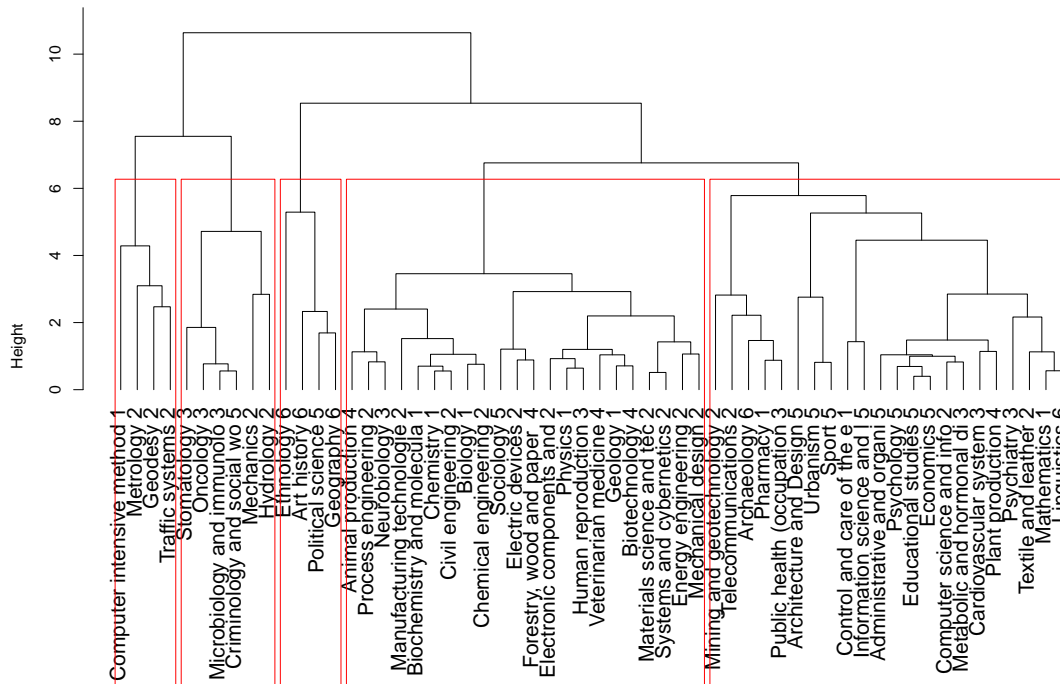


Fig. 2 Hierarchical clustering of the scientific disciplines

as natural and technological sciences. With one exception (Sociology), the disciplines come from the first four fields defined by the Slovene Research Agency in Table 6. This cluster is labeled NAT,ENG,MED,BIO in Table 5. The final cluster can be viewed only as a residual cluster with disciplines from all of the six fields in Table 2.¹³ It is labeled RESID.

| | trans. triads | same resGroup | deg alter | degOut | ExcelBeh |
|-------------------|---------------|---------------|-----------|--------|----------|
| 1 ENG,NAT | 0.12 | 0.48 | -0.23 | 0.30 | -0.25 |
| 2 ENG,MED | 0.19 | 0.29 | -0.53 | 0.13 | -0.01 |
| 3 HUM,SOC | 0.15 | 0.38 | -0.17 | -0.12 | 0.30 |
| 4 NAT,ENG,MED,BIO | 0.12 | 0.27 | -0.18 | 0.09 | -0.05 |
| 5 RESID | 0.07 | 0.18 | -0.08 | 0.05 | -0.00 |

Table 5 Averages of the importance coefficients according to the estimated parameters for each obtained cluster

The overall summary for the five clusters is shown in Table 5. The averages of the importance coefficients for each obtained cluster and each parameter are easy to summarize. First, for all clusters, the overall coefficients for transitive triads and belonging to the same research group are positive. The first of these results indicates the presence of a small-world clustering phenomenon and the second is for the impact of the institutional feature of belonging to the same research group. Second, the overall coefficients for the degree of alter is negative, contradicting the primary operationalization of preferential attachment, for all clusters. Third, the

¹³ Throughout our analyses, the seventh so-called field was ignored.

overall coefficients for all clusters except HUM,SOC are positive for the degree of alters outside the national disciplinary disciplines. This supports one aspect of the preferential attachment idea. Finally, these overall coefficients for the scientific excellence of alters are negative for all clusters except HUM,SOC. We note that these summary coefficients for the residual cluster follow all of these patterns but with the smallest values. These results are consistent with the results reported earlier and will not be summarized further.

Given this overall summary for the clusters, the next issue is whether the disciplines within the clusters have the same patterns of signs for the estimated parameters. Evidence for this assessment comes from the reported coefficients in Table 6. For all of the first four clusters, the coefficients for transitive triads and membership of the same research group follow the general pattern for *all* disciplines. Every discipline in the first cluster (ENG,NAT) has the same pattern of coefficients as for the whole cluster. In the second cluster (ENG,MED), this comes very close to holding completely regarding the coefficient's pattern for this cluster. Every discipline has the same pattern for the degree of alter and the degree outside the discipline. Regarding scientific excellence, we note that all estimated coefficients have the same sign as for the overall cluster.

For the third cluster (HUM,SOC), its disciplines have the same estimated coefficient pattern for degree of alter and scientific excellence as for the cluster as a whole. For degree outside the discipline, there is only one exception to the cluster's pattern. In the fourth cluster (NAT,ENG,MED,BIO), all of the disciplines in it have the same parameter sign for degree of alter. The same holds for degree outside the discipline but with only one exception out of 22 disciplines. Regarding scientific excellence, two disciplines have the wrong sign and three others have estimated coefficient values very close to 0. Overwhelmingly for the first four clusters of disciplines, the pattern of coefficient signs of the clusters are followed also by all the disciplines they contain. The overall summary regarding the substantive hypotheses is not driven by just a few disciplines: the phenomena hold at the disciplinary level.

Given the heterogenous nature of the RESID cluster, we cannot expect to see the same level of consistency. Even so, there is complete consistency for the disciplines regarding transitive triads. For membership in the same research group, there are only three inconsistencies (out of 22 disciplines) and these have negative values that are borderline. For the degree of alter, there are only four exceptions out of 22 disciplines. Even here, there is considerable consistency of the disciplines with the overall pattern. However, this is not the case for scientific excellence as half of the estimated coefficients have the wrong sign at the disciplinary level. We emphasize that small world clustering and preferential attachment phenomena hold at the disciplinary level even in the residual cluster.

6 Discussion and Conclusions

The first hypothesis about the presence of clustering as a dimension of a small-world structure was confirmed emphatically. The evidence regarding the second hypothesis concerning preferential attachment as the driving mechanism of co-authorship was decidedly mixed. Yet, in the main, our results contradict the hypothesis of a single preferential attachment mechanism for the formation of collaborative ties. The current number of co-authorships inside the field or discipline has a *negative* effect on new tie formation in all scientific fields and nearly all scientific disciplines. Our results show the distance between researchers who collaborate matters also. Regarding this, the social sciences and humanities differ from the other four fields. Alters higher degree of collaboration outside of Slovenia has a negative effect on new tie formation in the social sciences and humanities but a positive effect in the other fields. A high degree of collaboration outside the national disciplinary field, a revised notion of preferential attachment, has a positive impact on tie formation for the natural, technical, medical and biotechnical sciences. Clearly, preferential attachment mechanisms are more subtle than can be summarized by stating that researchers who currently have a large number of co-authors will see an increase in their number of collaborators, at least for the science dynamics in smaller national scientific systems.

Another difference between the social sciences and the humanities compared with the other four fields is that publication excellence had a positive effect on the formation of collaborative ties in the social sciences and humanities but a negative effect in the other fields. However, the standard hypothesis that young researchers form new co-authorship ties with scientifically excellent older scientists was not confirmed. Researchers from

all fields but the humanities were more likely to form new ties with younger colleagues. The third hypothesis was confirmed and the evidence demonstrates that the scientific fields and disciplines are affected by the organization of local institutions and publishing cultures. Overall, most of the results of Kronegger *et al.* (2012) were confirmed with the differences attributable to using better data containing more scientists, considering many disciplines and using a more complex and realistic SAOM specification. We note that our findings regarding the effects of small-world phenomena, preferential attachment, institutional arrangements hold for virtually all disciplines as well as for broad fields.

The differences between the two basic pools of scientific knowledge (i.e., between the natural and technical sciences, and the social sciences and humanities) according to the mechanism of preferential attachment are also the result of contextual (research policy) factors operating in Slovenia. In the former socialist era, due to ideological pressure on the social sciences and the humanities, these disciplines were less internationalized and much less oriented to publishing in high-ranking international journals. After the dramatic changes starting at the beginning of the 1990s, Slovenian R&D policy gradually began to introduce the criteria of (international) excellence into R&D evaluation procedures. In the first years after Slovenia's independence, governmental institutions created many initiatives, including legislation, to establish the new country including those relating to science. It seems that in adapting to the more demanding R&D evaluation criteria there were different impacts on the social sciences and humanities compared to the other fields. In part, this helps account for some of the differences between the two groups of fields. Yet, those social scientists and humanists who already engaged in excellent publication activity can be very attracted to research collaboration with their disciplinary colleagues. Generally speaking, the results of our bibliometric analysis showed that we cannot solely explain the differences in forming co-authorship ties in different parts of sciences with internal (epistemological) factors, but that also external (R&D policy) factors are involved. No single approach, be it sociological, social network analytic or based on conceptions drawn from physics, can be useful by itself. Multiple approaches are needed to account for the complex phenomenon of scientific collaboration, especially for national scientific systems. While science can be viewed as a general phenomenon it is also conditioned by local institutional contexts.

References

- Abbasi A, Hossain L, Leydesdorff L (2012) Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Infometrics* 6(3): 403–412
- Barabási AL (2002) *Linked: The New Science of Networks*, 1st edn. Basic Books
- Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509–512
- Çavuşoğlu A, Türker I (2013) Scientific collaboration network of Turkey. *Chaos solutions and Fractals* 57: 9–18, DOI 10.1016/j.chaos.2013.07.022
- avuolu A, Türker I (2014) Patterns of collaboration in four scientific disciplines of the Turkish collaboration network. *Physica A: Statistical Mechanics and its Applications* 413:220–229, DOI 10.1016/j.physa.2014.06.069
- De Stefano D, Giordano G, Vitale M (2011) Issues in the analysis of co-authorship networks. *Quality & Quantity* 45(5):1091–1107, DOI 10.1007/s11135-011-9493-2
- Ferligoj A, Kronegger L (2009) Clustering of Attribute and/or Relational Data. *Metodološki Zvezki (Advances in Methodology and Statistics)* 6(2):135–153
- Garfield E, Merton RK (1979) *Citation indexing: Its theory and application in science, technology, and humanities*. Wiley, New York
- Groboļjšek B, Ferligoj A, Mali F, Kronegger L, Igljč H (2014) The role and significance of scientific collaboration for the new emerging sciences: The case of Slovenia. *Teorija in praksa* 51(5):864–883
- Igljč H, Kronegger L, Ferligoj A, Doreian P (2014) *Forms and Sources of Research Collaboration: Evidence Across Four Scientific Disciplines*. Submitted
- Indlekofer N and Brandes U (2013) Relative importance of effects in stochastic actor-oriented models. *Network Science* 1(3):278–304

- 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
- Kronegger L, Mali F, Ferligoj A, Doreian P (2012) Collaboration structures in Slovenian scientific communities. *Scientometrics* 90(2):631–647, DOI 10.1007/s11192-011-0493-8
- Kronegger L, Mali F, Ferligoj A, Doreian P (2014) Classifying scientific disciplines in Slovenia: A study of the evolution of collaboration structures. *Journal of the Association for Information Science and Technology* pp n/a–n/a, DOI 10.1002/asi.23171
- Li L, Alderson D, Doyle JC, Willinger W (2006) Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications. *Internet Mathematics* 2(4): 431–523
- Mali F, Kronegger L, Ferligoj A (2010) Co-authorship trends and collaboration patterns in the Slovenian sociological community. *Corvinus Journal of Sociology and Social Policy (CJSSP)* 1(2):29–50
- Merton RK (1968) The Matthew Effect in Science. *Science* 159:56–63
- Merton RK (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 MEJ (2000) Small worlds: the structure of social networks. *Condmatt/0111070*
- Newman MEJ (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
- Perc M (2010) Growth and structure of Slovenia's scientific collaboration network. *Journal of Informetrics* 4:475–482, URL <http://arxiv.org/abs/1004.4824>
- Pike TW (2010) Collaboration networks and scientific impact among behavioral ecologists. *Behavioral Ecology* 21(2):431–435, DOI 10.1093/beheco/arp194
- Price DS (1963) *Little Science, Big Science and Beyond*. Columbia University Press, New York
- Price DS (1965) *Networks of Scientific Papers*. *Science* 149:510–515
- Price DS (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
- Said YH, Wegman EJ, Sharabati WK, 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
- Shrum W, Mullins N (1988) Network analysis in the study of science and technology. In: van Ran A (ed) *Handbook of Quantitative Studies of Science and Technology*, Elsevier, pp 107–133
- Shrum W, Genuth J, Chompalo I (2007) *Structures of scientific collaborations*. The MIT Press
- Snijders TA (2001) The statistical evaluation of social network dynamics. In: Sobel M, Becker M (eds) *Sociological Methodology—2001*, Basil Blackwell, pp 361–395
- Snijders TA (2005) Models for longitudinal network data. In: Carrington P, Scott J, Wasserman S (eds) *Models and Methods in Social Network Analysis*, Cambridge University Press, pp 215–247
- Snijders TA (2008) Analysing dynamics of non-directed social networks. Draft paper. URL http://www.stats.ox.ac.uk/~snijders/siena/TwoSided_s.pdf
- Snijders TA, Steglich C, Schweinberger M (2007) Modeling the co-evolution of networks and behavior. In: van Montfort K, Oud H, Satorra A (eds) *Longitudinal models in the behavioral and related sciences*, Mahwah, NJ: Lawrence Erlbaum, pp 41–71
- Snijders TA, van de Bunt GG, Steglich C (2010) Introduction to stochastic actor-based models for network dynamics. *Social Networks* 32(1):44–60
- de Sola Pool I, Kochen M (1978) Contacts and influence. *Social Networks* 1(1):5–51
- Wagner C, Leydesdorff L (2005) Network structure, self-organization, and the growth of interantional collaboration in science. *Research Policy* 34: 1608–1618
- Ward J (1963) Hierarchical grouping to optimize as objective function. *Journal of the American Statistical Association* 58:236–244
- Watts DJ, Strogatz SH (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 University Press., Cambridge

Acknowledgments

This research has been supported by the Slovenian Research Agency (10-ECRP-044).

Appendix

| numbers of disciplines and field name of discipline | rate (prevol 1) | rate (prevol 2) | degree (prevol) | transitive triads | same size group | degree of alter | detOut | ExciteIn | inPop | inPop similar | PPII | GrowthRate |
|---|-----------------|-----------------|-----------------|-------------------|-----------------|-----------------|--------|----------|--------|---------------|--------|------------|
| 1 Mathematics | 3.201 | 0.5113 | 6.690 | 0.869 | -1.211 | 0.1335 | 1.148 | 0.1335 | -0.143 | 0.0266 | -0.187 | 0.388 |
| 2 Physics | 11.721 | 0.6771 | 14.200 | 0.7133 | -1.425 | 0.0352 | 0.427 | 0.0352 | 0.156 | 0.042 | 0.269 | 0.168 |
| 3 Biology | 6.500 | 0.772 | 9.657 | 0.839 | -1.762 | 0.0082 | 0.957 | 0.0082 | 0.234 | 0.005 | 0.987 | 0.091 |
| 4 Chemistry | 13.199 | 0.802 | 12.454 | 0.720 | -1.467 | 0.0185 | 0.387 | 0.0185 | 0.152 | 0.025 | 0.184 | 0.125 |
| 5 Earth and planetary science | 13.893 | 1.502 | 19.155 | 2.144 | -1.243 | 0.081 | 0.553 | 0.081 | 0.182 | 0.083 | 0.172 | 0.102 |
| 6 Geology | 1.219 | 0.660 | 3.585 | 1.564 | -0.499 | 0.823 | 1.850 | 0.803 | 0.104 | 0.003 | -0.027 | 0.089 |
| 7 Computer intensive method | -1.199 | 0.809 | 1.211 | 0.826 | 0.525 | 0.435 | 0.155 | 0.435 | 0.557 | 0.074 | -0.730 | 0.311 |
| 8 Control and care of the e | 8.338 | 0.495 | 5.295 | 0.619 | -2.001 | 0.166 | 0.628 | 0.060 | 0.130 | 0.041 | 0.090 | 0.102 |
| 9 Civil engineering | 8.338 | 0.495 | 5.295 | 0.619 | -2.001 | 0.166 | 0.628 | 0.060 | 0.130 | 0.041 | 0.090 | 0.102 |
| 10 Chemical engineering | 8.338 | 0.495 | 5.295 | 0.619 | -2.001 | 0.166 | 0.628 | 0.060 | 0.130 | 0.041 | 0.090 | 0.102 |
| 11 Computer science and tec | 1.246 | 0.516 | 1.61 | 0.514 | -0.033 | 0.730 | 2.471 | 0.033 | -0.012 | 0.025 | -0.022 | 0.171 |
| 12 Mechanical engineering | 10.551 | 0.1086 | 9.998 | 0.753 | -1.574 | 0.107 | 0.497 | 0.027 | 0.176 | 0.097 | -0.266 | 0.204 |
| 13 Systems and cybernetics | 8.499 | 0.682 | 10.154 | 0.731 | -1.882 | 0.0085 | 0.619 | 0.031 | 0.041 | 0.041 | 0.074 | 0.088 |
| 14 Computer science and info | 6.665 | 0.208 | 8.440 | 1.382 | -1.656 | 0.308 | 0.913 | 0.079 | 0.080 | 0.080 | -0.181 | 0.105 |
| 15 Electronic components and | 3.975 | 0.544 | 6.796 | 0.977 | -2.131 | 0.163 | 0.639 | 0.056 | 0.090 | 0.161 | -0.280 | 0.281 |
| 16 Manufacturing technology | 3.975 | 0.544 | 6.796 | 0.977 | -2.131 | 0.163 | 0.639 | 0.056 | 0.090 | 0.161 | -0.280 | 0.281 |
| 17 Mechanical design | 3.064 | 0.600 | 5.748 | 0.879 | -1.700 | 0.217 | 0.716 | 0.065 | 0.081 | 0.081 | -0.307 | 0.171 |
| 18 Computer science and info | 2.657 | 0.682 | 4.861 | 1.084 | -1.079 | 0.405 | 0.857 | 0.185 | 0.141 | 0.141 | 0.196 | 0.327 |
| 19 Textile and leather | 4.244 | 0.720 | 4.042 | 0.916 | -1.200 | 0.195 | 0.680 | 0.100 | 0.226 | 0.111 | -0.404 | 0.284 |
| 20 Process engineering | 5.110 | 0.686 | 11.438 | 1.423 | -0.963 | 0.128 | 0.470 | 0.049 | -0.127 | 0.072 | 0.242 | 0.165 |
| 21 Metallurgy | 2.657 | 0.682 | 4.861 | 1.084 | -1.079 | 0.405 | 0.857 | 0.185 | 0.141 | 0.141 | 0.196 | 0.327 |
| 22 Textile and leather | 3.336 | 0.662 | 7.654 | 1.446 | -1.286 | 0.259 | 0.525 | 0.107 | 0.588 | 0.168 | -0.538 | 0.292 |
| 23 Genetics | 4.739 | 1.328 | 3.687 | 1.060 | -1.555 | 0.372 | 0.789 | 0.173 | 0.263 | 0.117 | -0.126 | 0.275 |
| 24 Traffic systems | 4.607 | 1.190 | 3.646 | 1.209 | -0.667 | 0.333 | 0.709 | 0.088 | 0.147 | 0.121 | 0.076 | 0.311 |
| 25 Hydraulics | 6.737 | 1.478 | 8.132 | 1.717 | -0.725 | 0.254 | 0.969 | 0.116 | 0.328 | 0.139 | -0.341 | 0.245 |
| 26 Neurobiology and immuno | 11.406 | 0.964 | 13.465 | 1.203 | -1.728 | 0.069 | 0.691 | 0.042 | 0.322 | 0.037 | -0.325 | 0.091 |
| 27 Oncology | 23.331 | 0.586 | 24.035 | 1.725 | -1.248 | 0.057 | 0.396 | 0.033 | 0.082 | 0.034 | 0.052 | 0.094 |
| 28 Immunology | 13.024 | 1.389 | 15.142 | 2.130 | -1.620 | 0.060 | 0.335 | 0.020 | 0.109 | 0.044 | 0.037 | 0.104 |
| 29 Cardiac and vascular system | 3.777 | 0.930 | 6.110 | 1.099 | -1.211 | 0.176 | 0.898 | 0.190 | 0.139 | 0.089 | -0.012 | 0.009 |
| 30 Metabolic and hormonal di | 5.637 | 0.876 | 7.832 | 1.182 | -2.067 | 0.149 | 0.841 | 0.097 | 0.034 | 0.062 | -0.248 | 0.151 |
| 31 Public health (occupation | 20.056 | 0.021 | 19.432 | 1.999 | -0.993 | 0.089 | 0.229 | 0.020 | 0.666 | 0.067 | -0.052 | 0.008 |
| 32 Forestry | 16.794 | 1.787 | 20.117 | 1.434 | -1.170 | 0.057 | 0.424 | 0.024 | 0.094 | 0.041 | -0.010 | 0.070 |
| 42 Animal production | 4.231 | 0.577 | 12.263 | 1.849 | -1.300 | 0.176 | 0.653 | 0.073 | 0.116 | 0.049 | 0.409 | 0.102 |
| 43 Plant production | 9.839 | 0.792 | 15.250 | 1.523 | -1.633 | 0.056 | 0.542 | 0.035 | 0.088 | 0.086 | -0.312 | 0.224 |
| 44 Biotechnological medicine | 16.979 | 3.400 | 17.196 | 1.843 | -1.785 | 0.059 | 0.688 | 0.040 | 0.041 | 0.021 | 0.001 | 0.021 |
| 45 Economics | 2.764 | 0.682 | 7.002 | 1.241 | -1.492 | 0.166 | 0.669 | 0.069 | -0.025 | 0.032 | 0.217 | 0.064 |
| 46 Social and behavioral sci | 9.408 | 0.802 | 7.002 | 1.241 | -1.492 | 0.166 | 0.669 | 0.069 | 0.104 | 0.146 | -0.313 | 0.064 |
| 47 Educational studies | 8.529 | 2.818 | 14.018 | 3.325 | -1.065 | 0.300 | 1.027 | 0.129 | 0.169 | 0.168 | 0.098 | 0.183 |
| 48 Economics | 5.617 | 0.437 | 2.887 | 1.200 | -0.994 | 0.134 | 0.562 | 0.039 | 0.134 | 0.041 | -0.011 | 0.009 |
| 49 Law | 4.922 | 1.064 | 9.050 | 1.499 | -1.375 | 0.060 | 0.488 | 0.030 | 0.045 | 0.029 | -0.044 | 0.065 |
| 50 Criminology and social wo | 1.910 | 1.017 | 4.487 | 1.017 | -0.838 | 0.520 | 0.338 | 0.045 | 0.369 | 0.023 | -0.012 | 0.704 |
| 51 Urbanism | 1.872 | 0.833 | 5.089 | 1.794 | -1.184 | 0.473 | 2.433 | 0.034 | 0.391 | 0.167 | -0.294 | 0.477 |
| 52 Architecture and Design | 5.617 | 0.437 | 2.887 | 1.200 | -0.994 | 0.134 | 0.562 | 0.039 | 0.045 | 0.029 | -0.044 | 0.065 |
| 53 Information science and I | 5.759 | 1.700 | 5.557 | 1.284 | -1.619 | 0.715 | 1.240 | 0.329 | 0.126 | 0.022 | -2.778 | 0.680 |
| 54 Linguistics | 3.757 | 0.794 | 5.557 | 1.282 | -1.484 | 0.185 | 1.950 | 0.249 | 0.106 | 0.057 | -0.006 | 0.005 |
| 55 Geography | 0.938 | 0.463 | 3.310 | 1.388 | -1.158 | 0.581 | 1.647 | 0.675 | 0.248 | 0.129 | -2.428 | 1.298 |
| 56 Art history | 1.713 | 1.460 | 9.480 | 1.253 | -1.723 | 0.204 | 0.939 | 0.160 | 0.096 | 0.059 | 0.083 | 0.257 |

Table 6 Estimated parameters for scientific disciplines according to the importance coefficients