

**This is manuscript.**

The final publ. is available at Springer via <http://dx.doi.org/10.1007/s11192-015-1790-4>.

## The stability of co-authorship structures

Marjan Cugmas · Anuška Ferligoj · Luka Kronegger

**Abstract** This article examines the structure of co-authorship networks' stability in time. The goal of the article is to analyse differences in the stability and size of groups of researchers that co-author with each other (core research groups) formed in disciplines from the natural and technical sciences on one hand and the social sciences and humanities on the other. The cores were obtained by a pre-specified blockmodeling procedure assuming a multi-core–semi-periphery–periphery structure. The stability of the obtained cores was measured with the Modified Adjusted Rand Index. The assumed structure was confirmed in all analysed disciplines. The average size of the cores obtained is higher in the second time period and the average core size is greater in the natural and technical sciences than in the social sciences and humanities. There are no differences in average core stability between the natural and technical sciences and the social sciences and humanities. However, if the stability of cores is defined by the splitting of cores and not also by the percentage of researchers who left the cores, the average stability of the cores is higher in disciplines from the scientific fields of Engineering sciences and technologies and Medical sciences than in disciplines of the Humanities, if controlling for the networks' and disciplines' characteristics. The analysis was performed on disciplinary co-authorship networks of Slovenian researchers in two time periods (1991–2000 and 2001–2010).

**Keywords** social networks · scientific collaboration · network dynamics · blockmodeling · Rand Index

---

M. Cugmas  
Faculty of Social Sciences, University of Ljubljana, Slovenia  
E-mail: [marjan.cugmas@fdv.uni-lj.si](mailto:marjan.cugmas@fdv.uni-lj.si)

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

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

This is manuscript.

Suggested citation:

Cugmas, Marjan, Anuška Ferligoj, and Luka Kronegger. 2016. The stability of co-authorship structures. *Scientometrics* 106 (1):163–186. doi: 10.1007/s11192-015-1790-4.

## 1 Introduction

Collaboration in science plays an important role in the production and dissemination of new scientific knowledge. (Beaver, 2004, pp. 399) presented strong and persuasive philosophical and scientometric arguments "in favor of collaborative research having greater epistemic authority than research performed by individual scientists alone". In order to achieve greater efficiency in the use of available resources, higher productivity, improved prestige and visibility, many R&D policies have been adopted to increase scientific collaboration at all levels and across all disciplines (Chinchilla-Rodríguez et al, 2012).

Many studies deal with scientific collaboration and several definitions of this phenomenon have been proposed. Laudel (1999, p. 32) defined scientific collaboration "as a system of research activities by several actors related in a functional way and coordinated to attain a research goal corresponding with these actors' research goals or interests". Later, he divided scientific collaboration into six categories: (i) collaboration involving a division of labour; (ii) service collaboration; (iii) the provision of access to research equipment; (iv) the transmission of know-how; (v) mutual stimulation; and (vi) trusted assessorship. Half of them are invisible in formal communication channels (Laudel, 2002). Katz and Martin (1997) argued that the borderline of scientific collaboration is unclear and that there is no accurate way to measure scientific collaboration.

Nevertheless, scientific collaboration is usually operationalised through co-authorship, which is treated as one of the main results of scientific collaboration. Yet this operationalisation is often criticised. Katz and Martin (1997) presented many cases of collaboration which do not result in co-authorship. Even within the various co-authorships there exists a different intensity of collaboration between the authors. Even if co-authorship is only one aspect,<sup>1</sup> of collaboration, it is still the most useful and efficient way of measuring research collaboration (Lundberg et al, 2006). Last but not least, a co-authorship publication represents one of the most formal manifestations of scientific communication (Groboljšek et al, 2014).

An approach commonly used for studying scientific collaboration through co-authorship is based on analysis of a co-authorship network. Here, the network units represent authors connected by an undirected line if they have co-authored one or more publications. In comparison with the often analysed citation network, where the network nodes are publications and the links between them are citations represented by asymmetric lines (Mali et al, 2010), co-authorship implies a much stronger social bond. Citations can occur without the authors knowing each other and therefore citation networks are not personal social networks. Further, citations span across time (Liu et al, 2005; Mali et al, 2012).

Co-authorship networks can be used to answer a broad variety of questions about collaboration patterns (Newman, 2004) and to investigate the correlation between certain authors' characteristics and the structure of their relationships. The results of many studies indicate a strong relationship between collaboration and the quality of the research as well as between collaboration and the speed of diffusion of scientific knowledge (Hollis, 2001; Frenken et al, 2005; Abbasi et al, 2011; Lee and Bozeman, 2005). Glänzel (2002) showed that "the extent of co-authorship and its relation with productivity and citation impact largely varies among fields". This relationship depends on how collaboration and productivity are measured. It also depends on the scientific discipline – the correlation between these phenomena can also be negative (Hu et al, 2014).

<sup>1</sup> De Haan et al (1994) distinguished five other indicators for measuring collaboration between researchers: shared editorship of publications; shared supervision in Ph.D. projects; writing research proposals together; participation in formal research programmes; and shared organisation of conferences.

The question about the way co-authorship ties are formed is often addressed in co-authorship network analysis. For example, Moody (2004) indicated that authors with many collaborators and high scientific prestige gain more connections from newcomers than their colleagues. Based on a study of a longitudinal co-authorship network of one scientific discipline,<sup>2</sup> Abbasi et al (2012) exposed the importance of node centrality to the attachment of new researchers to them. In contrast, less attention is given to the factors influencing the persistence of collaboration ties. This can be studied on the micro level (individuals), meso level (e.g., institutions) and macro level (e.g., countries). Beside individual-level factors, factors on the institutional level and national research policy have an impact on collaboration among scientists (Garg and Dwivedi, 2014).

Nevertheless, less attention has been paid to the factors influencing the persistence of collaboration ties. Many scholars have studied the benefits of cohesive social structures organised into well-defined, tightly knit communities of connected individuals (Lambiotte and Panzarasa, 2009). This can be partly extended on the level of long- and short-term scientific collaborations. Depending on the research area, long-term collaborations can lead to negative or positive effects on scientific work. As summarised by Lambiotte and Panzarasa (2009), long-term collaborations can result in greater stability of social processes, which can lead to long-term development on all research levels. Further, socially embedded links reduce competition and increase the motivation to transfer information. The easiest access to information leads to facilitating innovation, knowledge creation and scientific creative endeavours. Long-term collaboration enables a deeper insight into the research problem, but collaborations that last too long can also result in no or very slow progress. One reason can be the researchers' isolation as a result of a lack of connections with new social circles. On the other hand, there is strong pressure on specialists to extend their cooperation beyond their narrow disciplinary borders in the case of disciplines where new and fresh ideas are obsolete in a few years. Short-term collaborations are usually more dynamic, innovative and enable pooling of the knowledge of different specialists and, potentially, fields (Howells et al, 2003). In that case, the probability of further developing and extending the short-term collaboration is lower.

Kronegger et al (2011) studied the structure of co-authorship networks of four scientific disciplines (Physics, Mathematics, Biotechnology and Sociology) in four five-year periods (1986–2005). One of the key findings was that, regardless of the research discipline, the co-authorship structure very quickly consolidates into a multi-core–semi-periphery–periphery structure. The term core denotes the group of researchers who publish together in a systematic way with each other. The semi-periphery consists of researchers who co-author less systematically but have at least one co-authored publication with researchers from the discipline, while the periphery includes authors who publish just as a single author or with authors outside the boundary of the defined disciplinary network (Mali et al, 2012). Kronegger et al (2011) also addressed the question of the stability of cores (groups of researchers) in the context of studying the evolution of a blockmodel structure in time by a visual presentation.

The main goal of this article is to obtain the network structure of almost all scientific disciplines in Slovenia in two time periods (1990–2000 and 2001–2010). The stability of the obtained cores, measured by the Modified Adjusted Rand Index 1, for each scientific discipline in the two time periods is studied. The differences in core stability are explained by certain characteristics of the obtained blockmodels and networks.

---

<sup>2</sup> The analysis was done on bibliographic units written in the English language, which include the keyword *steel structures* and was published in the top 15 scientific journals from this field.

## 2 Research hypotheses

Ferligoj and Kronegger (2009) performed an analysis of the co-authorship<sup>3</sup> network of Slovenian sociologists who were registered at the Slovenian Research Agency (ARRS) in 2008 (95 researchers). They applied a blockmodeling approach to obtain the co-authorship structures of Slovenian sociologists. They concluded that there is a clear multi-core–semi-periphery–periphery structure, which is defined in the same way as described above (Mali et al, 2012), i.e., the core consists of a group of researchers who all co-author with each other in a more systematic way than with the others from a certain scientific discipline. There can be several cores on the diagonal of the blockmodel. The semi-periphery is defined by several researchers from the same scientific discipline who co-author with each other in a less systematic way. The periphery consists of researchers who do not co-author with any other researcher from the same scientific discipline, but they can co-author with researchers from a different scientific discipline or with researchers without a research ID assigned by the Slovenian Research Agency (SRA). Later, Ferligoj and Kronegger (2009) extended the analysis to four scientific disciplines in four time periods. They confirmed the multi-core–semi-periphery–periphery structure of the network and also pointed out that this structure is not present in newly emerging scientific disciplines in early time periods (Kronegger et al, 2011). Therefore, the first hypothesis is:

H1 Research disciplines consolidate into a multi-core–semi-periphery–periphery structure.

The distinction between the natural and technical sciences on one side and the social sciences and humanities on the other is quite common.<sup>4</sup> Kronegger et al (2011) studied the differences in co-authorship patterns of researchers from four scientific disciplines representing different fields. They highlighted that the differences among disciplines not only depend on the subject of research but also on the nature of the work. They confirmed that disciplines where teamwork is not so crucial for scientific activity have less co-authorship activity. This can be explained by the motives for scientific collaboration. Beaver and Rosen (1978) identified 18 of such motives. Some of them are similar to the motives for scientific collaboration studied by Melin (2000) on the micro-level using a survey and semi-standardised interviews. He found that collaboration is characterised by strong pragmatism and a high degree of self-organisation. His results show the importance of social reasons for collaboration (e.g., a long-standing relationship) and also of cognitive (e.g., research specialisation) and technical reasons (e.g., research equipment). The latter is especially important in the context of Big Science which started after World War II where scientists became even more dependent on large and very sophisticated instruments, while the existence of large research centres became closely connected with expensive research equipment that requires and integrates different technical expertise from various scientific fields (Groboljšek et al, 2014, p. 866–867). An increasing trend of collaboration (and co-authorship) is typical of many scientific disciplines (e.g., Borrett et al (2014), Kronegger et al (2015)), but its intensity varies across different scientific disciplines. Melin (2000) concluded that in the medical sciences there are almost always teams working together, from time to time collaborating with other teams, while in the humanities there are basically no teams and collaboration is

<sup>3</sup> The following co-authored publications were included: articles in journals with an impact factor, other original scientific articles, chapters in scientific monographs, and scientific monographs.

<sup>4</sup> The term *natural and technical sciences* in the current paper includes the following fields according to the ARRS Classification Scheme: Natural sciences and mathematics, Engineering sciences and technologies, Medical sciences, and Biotechnical sciences. The term *social sciences and humanities* includes the field of Social sciences and the field of the Humanities according to the ARRS Classification Scheme.

not very common. Kyvik (2003) also reports that the highest level of multi-authored papers in Norway (in the period 1980 to 2000) is in medicine, followed by the natural sciences, social sciences and humanities. Hu et al (2014) endorsed the hypothesis that collaborations in theoretical disciplines are less effective than those in experimental disciplines. Therefore, the following hypotheses will be tested:

- H2 The average size of cores is increasing in time.
- H3 The average size of all cores in the natural and technical sciences is greater than the average of all cores in the social sciences and humanities.
- H4 The stability of cores is lower in the social sciences and humanities than in the natural and technical sciences.

The stability of cores over time in a scientific discipline is defined by the following rules: (i) if any core does not change in time, the stability of a discipline does not decrease; (ii) if cores from the first time period merge in the second time period, the stability does not decrease; and (iii) if a core from the first time period splits into several cores or leaves the network in the second time period, the stability decreases.

### 3 Methods

The analysis has three main parts. In the first part, the multi-core–semi-periphery–periphery structure is tested by blockmodeling on most scientific disciplines (in the two time periods). In the second part, the stability of the cores thus obtained is measured by the Modified Adjusted Rand Index 1 for each analysed scientific discipline. In the last part, the stability of the cores obtained for each scientific discipline is explained by scientific fields using linear regression analysis. The hypothesis about the average size of the cores across disciplines is tested using a standard *t*-test.

#### 3.1 Blockmodeling

Some authors, e.g., Adams et al (2005), studied the (size of) research teams by counting the number of co-authors of a scientific paper. With blockmodeling, more comprehensive information about research teams, their integration into the co-authorship network and the general structure of the co-authorship network can be obtained.

The goal of blockmodeling is to reduce a large, potentially incoherent network to a smaller, comprehensible and interpretable structure. It is based on the idea that units in a network are grouped according to some meaningful equivalence (Batagelj et al, 1998). The most established approaches to equivalence are structural and regular equivalence (Doreian et al, 1994). Two units are structurally equivalent if they are connected to all others in the same way (Lorrain and White, 1971) and two units are regularly equivalent if they link in equivalent ways to other units that are also equivalent (White and Reitz, 1983). Every two units that are structurally equivalent are also regularly equivalent. In the case of structural equivalence, the blockmodeling procedure is more robust than the other equivalences – small changes in the network do not have a great impact on the final solution (Nooy et al, 2011).

The problem of blockmodeling can be seen as an optimisational clustering problem (Batagelj et al, 1992a). There are two basic approaches to blockmodeling (Batagelj et al, 1992b): indirect and direct. The indirect approach is generally used to determine the number of clusters, while the direct approach, performed by the procedure of local optimisation

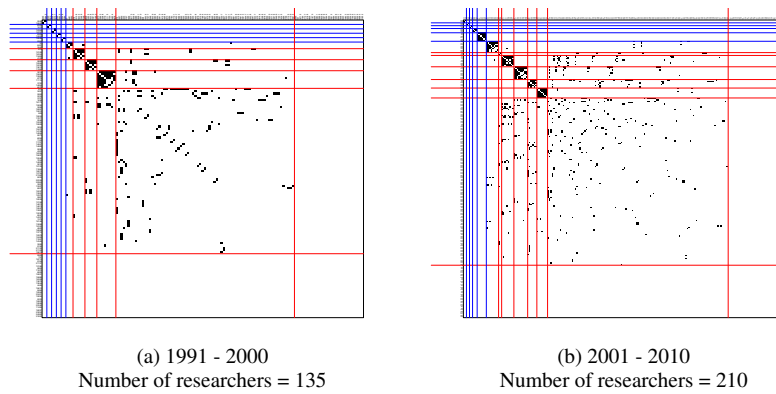


Fig. 1: Blockmodels of co-authorship networks of the discipline Mechanical design in two time periods

(Batagelj et al, 1992a), is used to produce the final blockmodeling solution. The algorithm implemented in Pajek (Nooy et al, 2011), a computer program for the analysis of a large networks, tries to decrease the criterion function by moving a randomly selected unit from one cluster to another or by interchanging two units in different clusters. The process continues until it can no longer improve the value of the criterion function (Nooy et al, 2011). The criterion function is the difference between the theoretical (ideal, assumed) and the empirical block structure.

Research by Kronegger et al (2011) on co-authorship networks showed that every structure of blockmodels became a multi-core-semi-periphery-periphery structure. In Figure 1, such a structure is denoted by a matrix where the rows and the columns represent the units (researchers) and the black dots in the cells in the matrix represent the (co-authorship) ties between the researchers.

As described above, in the case of the procedure of pre-specified blockmodeling of co-authorship networks the term core denotes the group of researchers from a scientific discipline (co-authorships outside this scientific discipline are not considered in this structure) who as co-authors published at least one scientific bibliographic unit together. In the first blockmodel in Figure 1, the nine cores are presented on the diagonal. Ideally, the cores are complete blocks with all possible connections (Mali et al, 2012).

The semi-periphery consists of researchers who co-author in a less systematic way (Mali et al, 2012). In the first blockmodel in Figure 1, the semi-periphery is the largest cluster with a few connections in its diagonal block. On the periphery there are authors who published just as a single author or with authors outside the boundary of the defined disciplinary network (Mali et al, 2012). In the first blockmodel in Figure 1, the periphery consists of the researchers in the last diagonal block without any connection.

The pre-specified blockmodeling procedure was performed using the Pajek program. As explained in the introduction, a multi-core-semi-periphery-periphery structure is assumed. The number of cores was estimated by observing the dendrogram obtained by the indirect approach where the corrected Euclidean distance (Burt et al, 1983) was used. The final blockmodel was obtained using the direct approach.

The algorithm implemented in Pajek (Batagelj et al, 1992a) had some difficulties detecting very small structural equivalent cores, especially in the case of those disciplines with a

very large number of researchers. To solve this problem, the following algorithm and Ward agglomerative procedure was used:

---

**Algorithm 1** Blockmodeling optimization algorithm

---

- 1: find the components that are also the cliques
  - 2: find the periphery
  - 3: extract (1) and (2) from the network
  - 4: run the blockmodeling procedure on the rest of the network
  - 5: merge the extracted part of the network with the solution obtained by the blockmodeling procedure
- 

The results of the blockmodeling procedure for the first and for the second time period are two partitions of researchers of a studied scientific discipline.

### 3.2 Measuring core stability

Many indices have been developed to measure the similarity of two partitions (Albatineh et al, 2006). Most of them assume a single set of units for classification, which may be impractical for the current research problem. In the case of studying the stability of cores of a co-authorship network, it has to be considered that some new researchers appear and others leave the network in the second time period. This means that the partitions are not obtained for the same set of units. Further, the splitting and merging of clusters have the same (negative) impact on the value of these measures.

Cugmas and Ferligoj (2015) proposed three versions of the Modified Rand Index which do not assume single set of units for classification, but there can be two, where one set is a subset of the other set. In the paper, one of the measures that they proposed, the Modified Rand Index 1 (MRI1), will be used. It assumes that the second set of units for classification is a subset of the first set of units. The data from the first and the second set of units are usually measured in two time points (periods). The term unit here denotes the researchers included in the blockmodeling. In the context of measuring the stability of the cores, the researchers present in the cores in the first time period but not present in the network in the second time period (e.g., retired researchers or those who did not co-author) or the researchers who were in one of the cores in the first time period but are in the semi-periphery or periphery in the second time period are denoted as drop-outs. Researchers not present in the first time period (newcomers) or researchers who were classified in the non-core part of the blockmodel were removed from the network as they do not affect the stability measure.

Given two sets of researchers  $S = \{O_1, \dots, O_n\}$  and  $T = \{O_1, \dots, O_m\}$ , where  $T \subset S$ , suppose  $U = \{u_1, \dots, u_r\}$  is the partition of  $S$  with  $r$  clusters and  $V = \{v_1, \dots, v_c\}$  is the partition of  $T$  with  $c$  clusters. The drop-outs  $S \setminus T$  define a new cluster  $v_{c+1}$  in partition  $V$ . When including  $v_{c+1}$  in  $V$ , the number of researchers in  $U$  is equal to the number of researchers in  $V$ . Let  $n_{ij}$  denote the number of researchers that are common to clusters  $u_i$  and  $v_j$ . Then the overlap between the two partitions can be written in the form of a contingency table where  $n_{i.}$  and  $n_{.j}$  are the numbers of units in clusters  $u_i$  (row  $i$ ) and  $v_j$  (column  $j$ ). If the cluster representing the drop-outs is arranged in the last column, then the contingency table is as follows:

Class	V					Sums
	$v_1$	$v_2$	...	$v_c$	$v_{c+1}$	
$u_1$	$n_{11}$	$n_{12}$	...	$n_{1c}$	$n_{1(c+1)}$	$n_{1.}$
.	.	.	...	.	.	.
.	.	.	...	.	.	.
.	.	.	...	.	.	.
$u_r$	$n_{r1}$	$n_{r2}$	...	$n_{rc}$	$n_{r(c+1)}$	$n_{r.}$
Sums	$n_{.1}$	$n_{.2}$	...	$n_{.c}$	$n_{.(c+1)}$	$n_{..} = n$

Then, the  $MRI1$  is defined as the proportion of all pairs of researchers in  $S \cap T$  that are placed in a cluster in  $U$  and in a cluster in  $V$  and all possible pairs of researchers of  $S$  that are placed in the same cluster in  $U$ :

$$MRI1 = \frac{\frac{1}{2} \sum_{i=1}^r \sum_{j=1}^c n_{ij}(n_{ij} - 1)}{\frac{1}{2} \sum_{i=1}^r \sum_{j=1}^{c+1} n_{ij}(n_{ij} - 1) + \frac{1}{2} (\sum_{i=1}^r n_i^2 - \sum_{i=1}^r \sum_{j=1}^{c+1} n_{ij}^2)} \quad (1)$$

The measure is defined on the interval between 0 and 1, where a higher value of the measure indicates a more stable classification. The value 1 is only possible when there are no drop-outs. The merging of clusters in the second time period does not result in a lower value of the measure, while splitting does. It is assumed that the merging of the two cores should not decrease the value of the measure of stability because merging indicates a higher number of ties (co-authorships) among researchers and this does not affect the ties created in the first time period. On the other hand, if the merging of the cores were to increase the value of the measure, the maximum value of the measure would be greater than 1.

Since the expected value of two random and independent partitions does not take a constant value, Cugmas and Ferligoj (2015) proposed an adjustment. The adjustment can be obtained by simulations in which the order of units of partition  $U$  and partition  $V$  is independently and randomly permuted many times. For each permutation of both partitions, the value of the  $MRI1$  is calculated. The mean of those obtained values represents the expected value of the  $MRI1$  in the case of two random and independent partitions. The value of the Modified Adjusted Rand Index 1 ( $MARI1$ ) is calculated as follows:

$$MARI1 = \frac{MRI1 - E(MRI1)}{1 - E(MRI1)} \quad (2)$$

The expected value of the  $MARI1$  in the case of two random and independent partitions is 0 and the maximum value is 1. As noted before, in the case of measuring the stability of cores of co-authorship blockmodels by the  $MRI1$ , the incoming researchers should be removed from the network. In the process of obtaining the expected value of the measure in the case of two random and independent partitions, this can be done before or after each permutation of the researchers of each partition. In the case of studying the stability of cores of co-authorship networks, the differences between the mentioned methods is expected to be very small (the value of Pearson's correlation coefficient on the empirical data is  $r = 0.99$ ). The differences can be explained by many factors. One of the most prominent is the percentage of researchers who were classified in the core part in the first time period and in the non-core part in the second time period.

In the current analysis, the researchers classified in the non-core part of the blockmodel were excluded from the network after the permutation procedure. In any case, the results of the (linear regression) analysis are almost equal, including the values of the regression coefficient and its statistical significance level.



#### 4 The data

The data sources are the Co-operative Online Bibliographic System and Services (COBISS) and the Slovenian Current Research Information System (SICRIS), maintained by the Institute of Information Science (IZUM) and ARRS.

SICRIS contains data about researchers (including information on their education, employment status, gender etc.) who have a research ID at the research agency. Information about the research groups, projects and research organisations to which the researchers belong is also gathered. COBISS is a national bibliographic database. Connecting these systems enables the formation of complete personal bibliographies of all researchers who have ever been registered at ARRS.

The analysis is performed on data for the period between 1991 and 2010. The data are aggregated into two consecutive ten-year time periods<sup>5</sup>, which reflect major changes in the R&D policies in Slovenia. The first time period, spanning from 1991 to 2000, is determined by the independence of the country, which meant that Slovenia started to adopt and implement its own science policies (Kronegger et al, 2011). In particular, R&D policy actors showed increasing interest in adopting its former classification system of science, which were affected by isolationism and parochialism before the independence of Slovenia, as well as in the standards of both the European Union and the Organisation for Economic Co-operation and Development (Kronegger et al, 2015). Moreover, the break in the trend in time series of the absolute number of published scientific publications and the development of the Internet and electronic communications is also characterised by the beginning of this time period Kronegger et al (2011). The second time period, from 2001 to 2010, is characterised by the country joining the European Union (EU) and adopting EU standards, and joining the North Atlantic Treaty Organization (NATO). In this time period, Slovenia started to participate in many EU and other international institutions. By the end of this time period, Slovenia had already (partly) integrated its national science system into the European one. In addition, ARRS was established in 2004, which had many positive effects on R&D evaluation procedures (Kronegger et al, 2015).

Collaboration between two researchers is operationalised by co-authorship of at least one relevant bibliographic unit according to ARRS: original, short or review articles, published scientific conference contributions, monographs or parts of monographs, scientific or documentary films, sound or video recordings, complete scientific databases, corpus and patents.

The analyses were conducted on the level of scientific disciplines according to the ARRS Classification Scheme, where six scientific fields<sup>6</sup> are defined: Natural sciences and mathematics (including 9 scientific disciplines), Engineering sciences and technologies (including 22 scientific disciplines), Medical sciences (including 9 scientific disciplines), Biotechnological sciences (including 6 scientific disciplines), Social sciences (including 13 scientific

---

<sup>5</sup> The analysis on four five years long time periods was also performed. The results regarding the block-model structure was the same as in the case of two ten-year long time periods.

<sup>6</sup> Actually, there are seven scientific disciplines in the ARRS Classification Scheme. The 7<sup>th</sup> is Interdisciplinary studies (including two disciplines: NCKS Research programme and Interdisciplinary research), which "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" (Ferligoj et al, 2015).

disciplines) and Humanities (including 12 scientific disciplines) (Table 1). In the analysis, 43 (out of 70) scientific disciplines were included<sup>7</sup>.

Most of the excluded disciplines were removed because of their small size in the first or second time period. Another reason for excluding a discipline was the absence of co-authorships in the current time period. One of the disciplines excluded is Theology that had not a single co-authored scientific bibliographic item published in the first time period.

The disciplines Computer science and informatics (565 researchers) and Chemistry (553 researchers) are the disciplines with the greatest number of researchers in the second time period. The number of researchers in the second time period is increasing in almost all analysed scientific disciplines (the average growth in the number of researchers in the second time period is 34 %). Exceptions are Veterinarian medicine (a 1.75 % decrease), Stomatology (a 1.54 % decrease) and Mining and geotechnology (a 10.00 % decrease) (Table 1).

The percentage of solo authors or authors who publish only in co-authorship with authors outside the discipline (the periphery) is decreasing in time. The average share of researchers in the periphery in the first time period is 38.7 %, while in the second time period it is 29.9 %. The largest decrease in the percentage of the periphery in the second time period is observed in the discipline Criminology and social work (a 64.6 % decrease).

The analyses dealing with scientific fields can be less accurate because of the specific national classification scheme that classifies research disciplines into research fields. For example, the discipline Geography is classified into the field of the Humanities according to the ARRS Classification Scheme, while according to the Common European Research Classification Scheme (CERIF) the discipline Geology, physical geography is classified in the field of the Natural sciences and mathematics. Some studies also show differences in the publication culture across scientific disciplines (Kronegger et al, 2015). According to the latter, the discipline Geography is more similar to the natural and technical fields than to the Humanities (Kronegger et al, 2015).

Another key characteristic on the level of scientific disciplines is the existence of sub-disciplines within scientific disciplines (according to De Haan et al (1994) this reflects the formation of research groups due to the cultural capital of researchers). The best known example is Physics which can be divided into theoretical and experimental physics. Such a distinction can be made in almost all scientific disciplines and is reflected in different ways, e.g., Yoshikane et al (2006) identified differences between the leader and the follower in the discipline of computer sciences (based on co-authorship). They concluded that the mentioned rules are separated more clearly in the application area than in the theoretical area. Similarly, Moody (2004) concluded that quantitative work is more likely to be co-authored than non-quantitative work in Sociology.

## 5 Results

The results of the described analysis will be presented in two parts. First, visualisations of the transitions will be shown along with a general and detailed interpretation of the measure

<sup>7</sup> Due to their small size (according to the number of researchers in the first or second time period), the following disciplines were excluded from the analysis: Computer-intensive methods and applications, Control and care of the environment, Geodesy, Mechanics, Traffic systems, Mining and geotechnology, Hydrology, Technology-driven physics, Communications technology, Psychiatry, Stomatology, Landscape design, Architecture and design, Information science and librarianship, Criminology and social work, Sport, Ethnic studies, Theology, Urbanism, Anthropology, Archaeology, Ethnology, Philosophy, Culturology, Literary sciences, Musicology and Art history.

Table 1: Considered scientific disciplines and their characteristics in two time periods

Scientific discipline and scientific field	1991 - 2000			2001 - 2010			average core size	periphery (%)	average core size	% of researchers that left the cores
	N	number of cores	semi-periphery (%)	N	number of cores	semi-periphery (%)				
1. Natural sciences and mathematics										
Biochemistry and molecular biology	171	11	45.61	30.41	301	17	45.85	33.55	4.13	68.29
Biology	279	12	46.59	37.63	427	18	55.04	27.40	4.69	70.45
Chemistry	443	22	64.79	12.87	553	29	67.09	11.93	4.30	68.09
Geology	211	14	44.55	34.12	294	10	43.54	42.86	5.00	64.44
Mathematics	155	9	34.19	51.61	257	13	51.36	32.30	3.82	59.09
Pharmacy	228	15	51.75	25.00	285	8	71.58	14.04	6.83	77.36
Physics	307	15	61.89	15.96	451	15	65.63	12.42	7.62	52.94
2. Engineering sciences and technologies										
Chemical engineering	153	8	60.13	24.18	197	10	66.50	18.78	3.62	62.50
Civil engineering	242	15	65.29	15.70	289	18	61.94	13.49	4.44	69.57
Computer science and informatics	401	25	51.87	27.43	565	21	63.36	21.24	4.58	62.65
Electric devices	70	8	31.43	25.71	97	6	38.14	26.80	8.50	56.67
Electronic components and technologies	116	14	36.21	23.28	135	15	40.74	25.93	3.46	61.70
Energy engineering	210	15	57.14	16.67	283	14	60.07	15.90	5.67	70.91
Manufacturing technologies and systems	217	14	45.62	28.57	276	14	52.90	23.55	5.42	50.00
Materials science and technology	253	13	60.08	17.39	332	14	62.05	13.86	6.15	61.40
Mechanical design	135	11	55.56	21.48	210	13	56.19	17.62	5.00	64.52
Metrology	74	12	39.19	20.27	100	10	35.00	30.00	3.89	56.67
Process engineering	92	11	52.17	11.96	111	11	48.39	17.20	3.56	66.67
Systems and cybernetics	240	12	54.17	21.67	291	14	58.42	17.53	5.83	48.28
Telecommunications	142	15	39.44	32.39	195	13	45.13	21.03	6.00	50.00
Textile and leather	100	11	55.00	12.00	131	9	61.07	11.45	5.14	63.64
3. Medical sciences										
Cardiovascular system	224	11	54.91	16.52	268	8	67.54	12.69	8.83	57.81
Human reproduction	164	7	53.66	21.95	187	7	55.61	11.76	12.20	42.50
Metabolic and hormonal disorders	74	12	13.51	45.95	97	11	38.14	28.87	3.56	66.67
Microbiology and immunology	208	10	54.81	14.90	270	8	67.78	8.89	10.50	47.62
Neurobiology	296	22	51.01	25.68	366	12	69.67	16.67	5.00	81.16
Oncology	233	10	54.08	17.60	277	11	56.32	13.36	9.33	45.45
Public health (occupational safety)	195	17	34.87	41.54	264	12	56.06	29.55	3.80	80.43
4. Biotechnical sciences										
Animal production	127	11	64.57	6.30	158	12	48.73	7.59	6.90	35.14
Biotechnology	111	9	49.55	18.92	196	9	58.16	13.78	7.86	60.00
Forestry, wood and paper technology	149	12	62.42	11.41	153	10	53.59	11.76	6.62	43.59
Plant production	247	10	68.83	12.96	312	11	66.35	8.33	8.78	35.56
Veterinarian medicine	171	10	60.23	8.19	168	8	61.31	5.36	9.33	51.85
5. Social sciences										
Administrative and organisational sciences	96	6	17.71	64.38	155	16	34.19	41.94	2.64	58.82
Economics	338	20	49.11	32.84	477	22	61.01	21.17	4.25	73.77
Educational studies	303	19	38.28	39.60	404	17	51.24	33.91	4.00	76.12
Law	157	10	24.20	47.13	204	15	31.37	35.78	5.15	40.00
Political science	98	12	23.47	48.98	142	9	45.77	31.69	4.57	74.07
Psychology	92	9	29.35	48.91	131	10	38.17	36.64	4.12	65.00
Sociology	135	11	28.15	42.22	203	9	48.77	31.03	5.86	45.00
6. Humanities										
Geography	87	8	33.33	35.63	117	8	53.85	24.79	4.17	40.74
Historiography	186	11	9.14	74.73	251	11	23.11	60.56	4.56	43.33
Linguistics	196	15	10.71	71.94	296	25	22.64	55.07	2.87	61.76

This is manuscript.

of stability. In the second part, the hypotheses about the average core size will be tested using a general  $t$ -test and the hypothesis about the stability of the cores across scientific fields will be tested using a linear regression model.

### 5.1 The structure of the obtained blockmodels and core stability

Figure 2 visualises the transitions of researchers in the two time periods for each analysed discipline. Each plot has two parts. The black rectangles on the top part of each plot represent the cores in the first time period and the black rectangles on the bottom represent the cores in the second time period. The rectangles located third from the right (orange coloured) represent the semi-periphery while those second from the right (red coloured) represent the periphery. The researchers who were not present in the network in the first time period are added to the partition obtained for the first time period. These researchers are called newcomers and are represented by the last rectangle on the top (green coloured) of each visualisation. Similarly, the researchers who left the network in the second time period are added to the partition obtained for the second time period. These researchers are called drop-outs and are represented by the last (green coloured) rectangle on the bottom of each visualisation. The number of researchers is then the same in both time periods. The edges connecting the clusters in partitions in the first time period and in the second time period represent the transition of the researchers between the clusters of two partitions.

The visualisations of the transitions confirm the first hypothesis about the co-authorship structure as well as the assumption that the majority of researchers' transitions happen in the non-core part of the blockmodel (also mentioned in the previous section). This is the common characteristic of all analysed scientific disciplines. The majority (between 58 % and 86 % in both time periods) of researchers were classified in the non-core part of the blockmodels, including newcomers. Many of them were classified into the semi-periphery or periphery. Abbasi et al (2012) obtained similar results in the field of *steel structures*. They observed many new authors who were not connected to any of the previously existing authors. While the number of researchers is larger in the second time period in almost all analysed disciplines, the size of the non-core part cannot be directly interpreted by the plot of transitions in Figure 2 because of the groups of newcomers and drop-outs.

In general, the core stability of the analysed disciplines in time is relatively low (Figure 2). There are all kinds of transitions between the obtained cores. Some researchers moved from the core to the semi-periphery or periphery, some stayed in the core and others even disappeared from the network in the second time period. There are many cores that moved into the non-core part of the blockmodel, some cores split into two or more new cores, and some merged into one core. These are all the possible transitions of researchers between the two time periods, which affect the value of the proposed measure of the stability of the cores in the two time periods.

The cores obtained by the pre-specified blockmodeling procedure are least stable in the discipline of Biochemistry and molecular biology ( $MARI1 = 0.01$ ). A relatively high percentage of cores moved to the semi-periphery or to the periphery in the second time period. Some researchers also left the co-authorship network in the second time period (Figure 3b).

The discipline with the highest value of  $MARI1$  is Plant production ( $MARI1 = 0.49$ ). Some relatively small cores moved to the semi-periphery or left the co-authorship network in the second time period, but not one researcher moved from the cores into the periphery. The cores with the highest number of researchers stay relatively stable in the second time

This is manuscript.

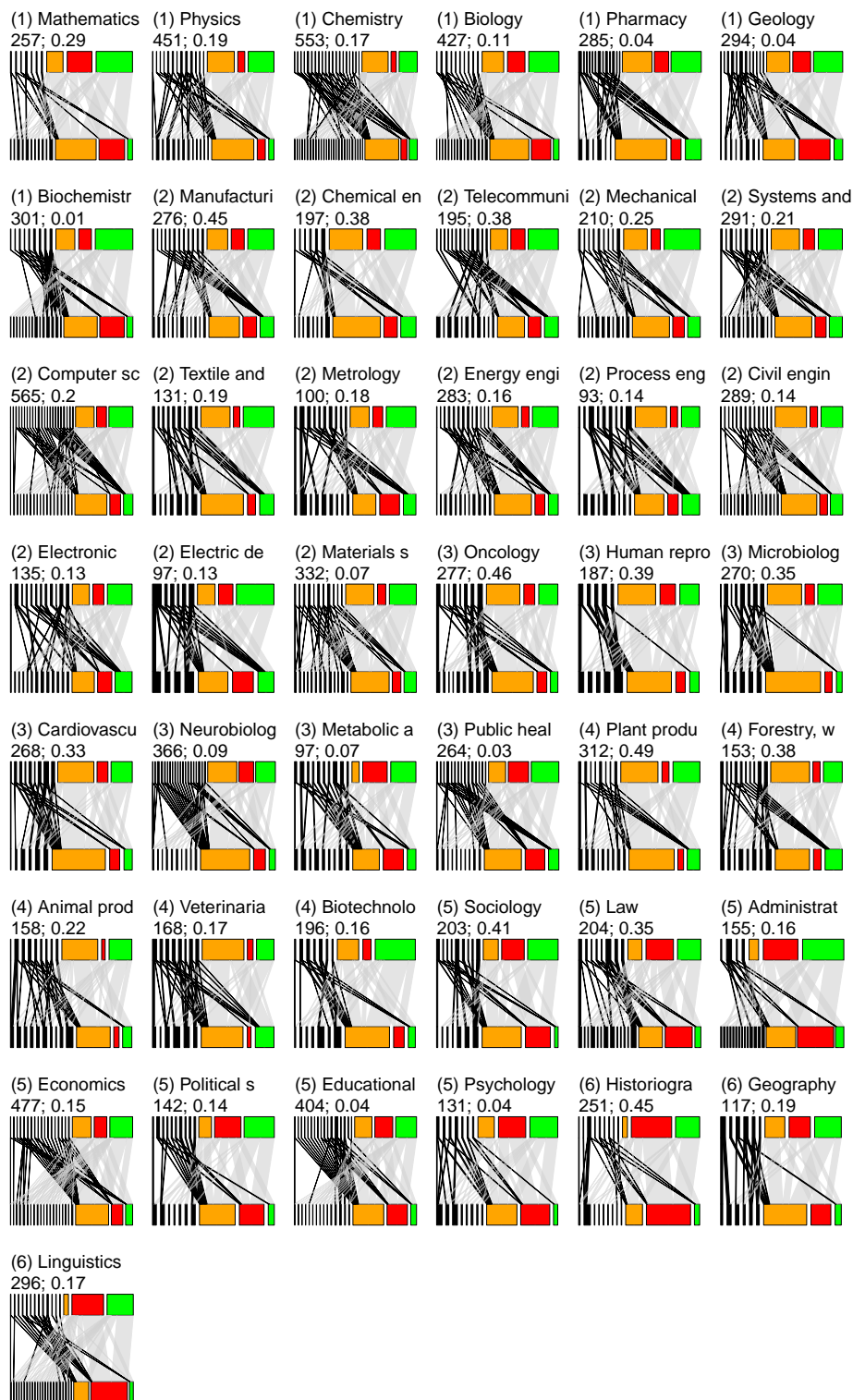


Fig. 2: The transitions of researchers in the two time periods (red the number in parantheses denotes the scientific field, the values under the name of a discipline present the number of researchers in the first time period and the value of MARI)

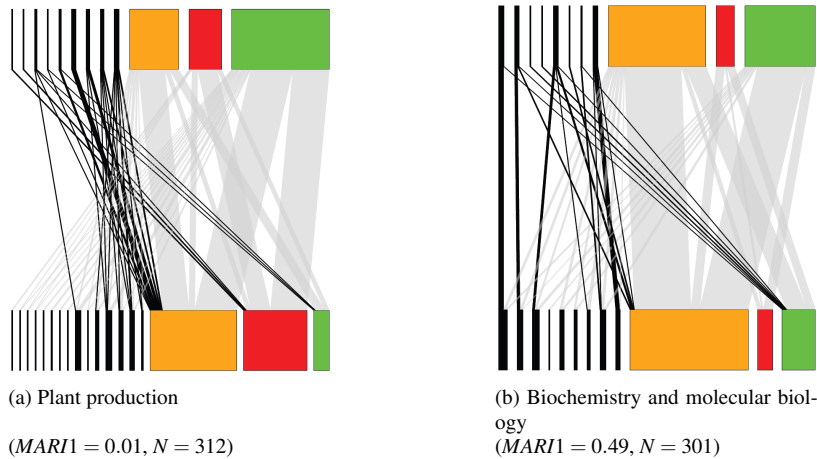


Fig. 3: The discipline with the least stable cores and the discipline with the most stable cores in the two time periods

period as well. There are also some splits that decrease the value of the index and also some merged ones which do not decrease the value of the index (Figure 3a).

## 5.2 Core stability: the natural and technical sciences vs. the social sciences and humanities

Following the well-known distinction between the natural and technical sciences and the social sciences and the humanities, the comparison of the average stability (measured by MARI1) of the cores can be treated by the dichotomous variable: *the natural and technical sciences* and *the social sciences and humanities*. For this purpose, the scientific fields are classified into two categories: fields 1, 2, 3, 4 into the category *the natural and technical sciences* and 5, 6 into the category *the social sciences and humanities*.

The hypothesis about the higher average stability of cores in the disciplines classified into the natural and technical sciences ( $MARI1 = 0.21$ ), compared to the average stability of cores of disciplines classified into the social sciences and humanities ( $MARI1 = 0.21$ ), could not be rejected ( $t = -0.06, df = 30, 15, p > 0.95$ ). As there is a high level of variability in the characteristics of the co-authorship networks and the blockmodel structures across scientific disciplines, the relationship across scientific fields and core stability was controlled by these characteristics in the linear regression model. In this section, the units are scientific disciplines ( $n = 43$ ). The dependent variable in the regression model is the stability of cores measured by the Modified Adjusted Rand Index 1 (MARI1). The considered explanatory variables are scientific fields (the natural and technical sciences vs. the social sciences and humanities) and selected control variables which are some characteristics of the networks and characteristics of the obtained blockmodels. Some variables were measured in the first time period and some in the second time point. A detailed description of the control variables describing the networks' characteristics and blockmodels' characteristics is:

- **the networks' characteristics;**
  - **the number of researchers** (1st time point) and **the growth of the number of researchers** (1st and 2nd time point);

The latter variable to some extent indicates the saturation of a specific discipline. According to Price (1963), the (growth of) scientific production follows a logistic S-curve. After the exponent phase, the linear phase follows and then saturation.

- **growth of density** (1st and 2nd time point) and **average core size** (1st time point); As described in the previous section, the density presents the share of all realised ties from all possible ties. This value often depends on the number of researchers and cannot be interpreted as an indicator of structural cohesion, especially because of the many subgroups in the networks (Friedkin, 1981). However, the value is typically greater in the case of smaller networks with a low rate of periphery and cores with many researchers. Controlling for other network characteristics, the following can be argued: in the case of a greater density, there are more researchers who co-authored only occasionally (semi-periphery) and more complete cores with a higher number of researchers. Therefore, the probability of creating ties with new researchers is lower and the stability of the cores is higher.
- **the blockmodels' characteristics;**
  - **percentage of cores** (average percentage of the 1st and 2nd time point); The variable is defined as the percentage of the number of researchers classified into the cores compared to all researchers in a certain scientific discipline. Controlling by the other explanatory variables, the negative impact of the average percentage of cores on the stability of cores may be explained by follows: in scientific disciplines where working in research teams in a systematic way is not so common, those who still work in that way are more probably part of long existing and well-established research groups whose work requires some special research equipment or they are part of a research group which is involved in any other long-term projects.
  - **presence of the bridge** (1st time point); The bridging core is defined as a group of researchers (core) who also publish together with at least two other groups of researchers (cores) in a systematic way. They connect two or more groups of researchers. This kind of collaboration can result in the merging of two or more cores in the latter time period.

As can be seen from Table 2, there is no statistically significant difference in average core stability ( $p < 0.05$ ) between the natural and technical sciences and the social sciences and humanities, controlling for other explanatory variables. With all explanatory variables included, 20 % of the variance (*Adjusted R Square*) of *MARI1* can be explained. It can also be seen that the average core size and growth of density have a statistically significant impact on the stability of the cores ( $p < 0.005$ ). The impact of both explanatory variables is positive. As described earlier, when controlling for other network and blockmodel characteristics the greater density indicates that there are many complete cores with a higher number of researchers in the discipline. The probability of creating new ties with other researchers is then lower and the stability of the cores is higher. Similarly, De Haan et al (1994) mentioned that the size of the research group affects the persistence of collaboration.

According to the already mentioned differences between the ARRS Classification Scheme and the publication cultures across various disciplines, further analysis is performed on the level of all six fields rather than the level of two classes of fields. The Humanities is used as the reference field since many studies suggest that the Social sciences are becoming more similar to the natural and technical sciences with regard to their publishing behaviour (Kyvik, 2003; Kronegger et al, 2015). The results (Table 3) are similar to the analysis on the level of two classes of scientific fields presented in Table 2. The considered explanatory variables explain 23 % of the variability of the stability of the cores. The average core

Table 2: The impact of the characteristics of the network, blockmodel and two classes of scientific disciplines on the stability of the cores

	<i>b</i>	<i>SE(b)</i>	<i>p</i>
intercept	0.0681	0.1936	0.73
number of researchers (first time point)	-0.0002	0.0003	0.46
growth of number of researchers (1st and 2nd time point)	0.0000	0.0014	0.96
growth of density (1st and 2nd time point)	0.0015	0.0006	0.03
average core size (1st time point)	0.0553	0.0154	0.00
percentage of cores (1st and 2nd time point)	-0.0028	0.0046	0.55
presence of the bridge (1st time point)	0.0266	0.0420	0.59
Natural and technical sciences vs. social sciences and humanities	-0.0276	0.0493	0.65
<i>Number of observations (disciplines):</i>			43
<i>R<sup>2</sup>:</i>			0.34
<i>Adjusted R<sup>2</sup>:</i>			0.20
<i>Residual std. error:</i>		0.12 (df = 35)	
<i>F Statistics:</i>	2.521 (df = 7 and 35)	(p < 0.03)	
<i>Method of estimation:</i>	Least Squares Method		

stability of scientific disciplines classified in the scientific field of the Humanities is higher than with other scientific fields. However, when controlling for other variables the differences between the Humanities and all other scientific fields are not statistically significant ( $p < 0.05$ ). There are two control variables with a statistically significant impact ( $p < 0.05$ ) on the stability of the cores as in the case of two classes of disciplines (growth of density and average core size).

Table 3: The impact of the characteristics of the network, blockmodel and disciplines on the stability of the cores

	<i>b</i>	<i>SE(b)</i>	<i>p</i>
intercept	0.0906	0.2027	0.66
number of researchers (first time point)	-0.0002	0.0003	0.58
growth of number of researchers (1st and 2nd time point)	0.0010	0.0015	0.53
growth of density (1st and 2nd time point)	0.0015	0.0010	0.04
average core size (1st time point)	0.0625	0.0177	0.00
percentage of cores (1st and 2nd time point)	-0.0054	0.0049	0.28
presence of the bridge (1st time point)	0.0404	0.0450	0.38
<b>Humanities (reference category)</b>			
Natural science and mathematics	-0.1511	0.0892	0.10
Engineering sciences and technologies	-0.0120	0.0834	0.89
Medical sciences	-0.0850	0.0954	0.38
Biotechnical sciences	-0.0353	0.1008	0.72
Social sciences	-0.0707	0.0844	0.41
<i>Number of observations (disciplines):</i>			43
<i>R<sup>2</sup>:</i>			0.43
<i>Adjusted R<sup>2</sup>:</i>			0.23
<i>Residual std. error:</i>		0.12 (df = 31)	
<i>F Statistics:</i>	2.151 (df = 11; 31)	(p < 0.05)	
<i>Method of estimation:</i>	Least Squares Method		

The stability of cores is operationalised by four main rules (see Section 2). In summary, the splitting of cores and the percentage of researchers who leave the cores are the main factors which are defined as those that reduce the stability of cores. By including in the



linear model the percentage of drop-outs as an explanatory variable, the stability of the cores can only be interpreted by the splitting of the cores due to the fact that the percentage of drop-outs is part of the core stability index (MARI1). Significant differences between the Humanities and other scientific fields would indicate that the different stability of cores across scientific fields is caused by the splitting of the cores.

Table 4: The impact of the characteristics of the network, blockmodel, disciplines and percentage of drop-outs on the stability of the cores

	<i>b</i>	<i>SE(b)</i>	<i>p</i>
intercept	0.8349	0.1840	0.00
number of researchers (first time point)	0.0001	0.0002	0.77
growth of number of researchers (1st and 2nd time point)	0.0004	0.0010	0.72
growth of density (1st and 2nd time point)	0.0091	0.0005	0.07
average core size (1st time point)	0.0053	0.0152	0.73
percentage of cores (1st and 2nd time point)	-0.0069	0.0033	0.05
presence of the bridge (1st time point)	-0.0005	0.0313	0.99
percentage of drop-outs	-1.0160	0.1667	0.00
Humanities (reference category)			
Natural science and mathematics	0.0378	0.0680	0.58
Engineering sciences and technologies	0.1339	0.0615	0.04
Medical sciences	0.1421	0.0748	0.07
Biotechnical sciences	0.0338	0.0694	0.63
Social sciences	0.0847	0.0626	0.19
<i>Number of observations (disciplines):</i>			43
<i>R<sup>2</sup>:</i>			0.75
<i>Adjusted R<sup>2</sup>:</i>			0.65
<i>Residual std. error:</i>			0.081 ( <i>df</i> = 30)
<i>F Statistics:</i>			7.375 ( <i>df</i> = 12; 30) ( <i>p</i> < 0.01)
<i>Method of estimation:</i>			Least Squares Method

The average core stability is lowest in the field of the Humanities. The average value is highest in Engineering and technologies ( $p < 0.05$ ) and Medical sciences ( $p < 0.10$ ). The values of the regression coefficients indicate that the average stability of the Natural sciences and mathematics and Biotechnological sciences is greater in comparison to the Humanities, but the differences are not statistically significant (Table 4).

The growth of density ( $p < 0.10$ ), the average percentage of cores ( $p < 0.05$ ), and the percentage of drop-outs ( $p < 0.01$ ) have statistically significant impacts on the core stability. The impact of the average core size on the core stability is not statistically significant ( $p < 0.10$ ) when the variable the percentage of drop-outs is included in the model as an explanatory variable (Table 4). Most of the explained variance of MARI1 (the current model explains 65 % of variance of MARI1) is caused by this variable, indicating that the instability of the cores is a consequence of short-term collaborations and less because of the splitting of the cores.

### 5.3 The difference in average core size between the natural and technical sciences and the humanities and social sciences

Many studies have observed an increasing trend in co-authorship and scientific collaboration. As stated, each analysed scientific discipline consists of periphery, semi-periphery and

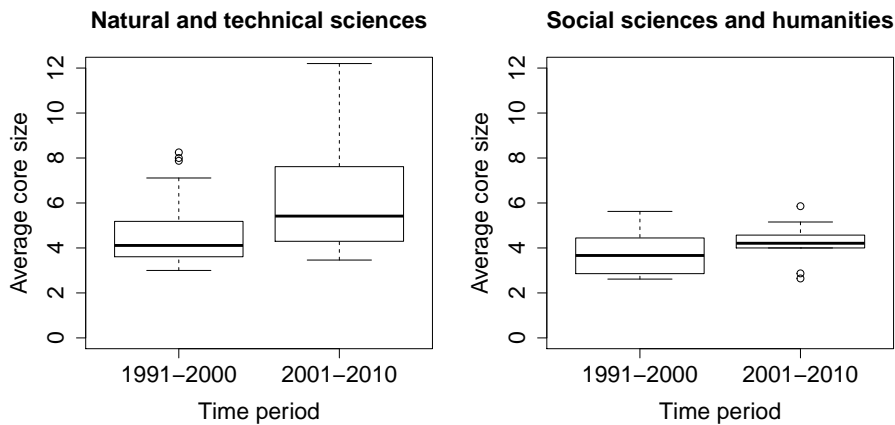


Fig. 4: The average core size by field and time period

several cores. There can be less or more cores of a different size, e.g., the discipline Administrative and organisational sciences consist of 6 cores in the first time period and 16 cores in the second time period. While there are more clusters in the second time period, the average number of researchers classified in each cluster is lower (the average core size in the first time period is 4.25, while the average core size in the second time period is 2.64) (Table 1). From Table 1, the average core size for each time period can be also calculated. The overall average core size in the first time period is 4.4 researchers, while in the second time period the overall average core size is 5.6 researchers ( $t = -3.03$ ,  $df = 72.35$ ,  $p < 0.01$ ). Across the disciplines, the highest average core size in the first time period is observed with Oncology (8.3 researchers) and Human reproduction (8.0 researchers), while the lowest average core size in the first time period is observed with Linguistics (2.6 researchers) and Psychology (2.9 researchers). Following the distinction between the natural and technical sciences and the social sciences and humanities, the comparison of the average size of the cores can be treated by the dichotomous variable. The average size of the cores in the fields of the natural and technical sciences is 4.6 researchers, while in the fields of the social sciences and humanities the average size of the cores is 3.8 researchers ( $t = -2.18$ ,  $df = 23.41$ ,  $p < 0.05$ ) (Figure 4).

## 6 Conclusions

Many studies have sought to explain the factors that promote collaboration in science, where the collaboration is usually operationalised by co-authorship. This article addresses the stability of the cores obtained by pre-specified blockmodeling on networks measuring co-authorship in the periods 1991–2000 and 2001–2010 in the Slovenian science system. The obtained cores represent those groups of researchers who systematically co-authored. The stability of the obtained cores is measured by the Modified Adjusted Rand Index 1 (MARI1).

Four main hypotheses were tested. The first hypothesis addressed the structure of the co-authorship network, the second and third the average size of the cores, while the fourth

considered the average stability of the obtained cores across disciplines by fields in the two time periods.

Findings about the structure of co-authorship networks proposed by previous research studies (Ferligoj and Kronegger, 2009; Kronegger et al, 2011) were confirmed. Each analysed scientific discipline has a multi-core–semi-periphery–periphery structure. The majority of researchers were classified into the semi-periphery or periphery in all of the analysed scientific disciplines. Between 14 % and 44 % of all researchers in the blockmodel (in both time periods) were classified into the cores in the blockmodels. However, as observed by many authors, the average number of authors who publish as co-authors is increasing in time. Despite the differences among the disciplines within scientific fields and even the differences within the scientific disciplines (such as the frequency of collaboration with foreign authors which were not included in the analysis), the differences in the average core size between the natural and technical sciences and the social sciences and humanities are statistically significant. The average core size is higher in the natural and technical sciences.

The differences in the average size of cores can be affected by the fact that authors from abroad are not included in the analysis. This would be reflected in cores with a lower number of authors in the natural and technical sciences. Kronegger et al (2015) reported that the rate of co-authored publications with researchers from abroad is higher in the fields of the natural and technical sciences than in the social sciences and humanities. Kyvik (2003) listed three main reasons for differences in the rate of publications in English and domestic languages across the mentioned fields. The listed reasons can be generalised on the level of scientific collaboration: (i) the tendency to study phenomena that have a specific geographic and social context; (ii) the poorer publishing possibilities in international or English-language journals; and (iii) the less motivating reward system for international publishing in Sociology.

The hypothesis about the stability of the obtained cores in time was partly confirmed. There are no statistically significant differences in the average stability of cores between the natural and technical sciences and the social sciences and humanities, even when considering some control variables (number of researchers in the scientific discipline, growth in number of researchers, growth in density, average core size, average percentage of cores, presence of a bridge). The results are also similar when the six scientific fields are included in the linear model instead of two categories represented by the natural and technical sciences on one hand and the social sciences and humanities on the other.

If the stability of cores is defined only by the splitting of cores and not by the percentage of drop-outs, there are some statistically significant differences in core stability across the considered scientific fields' impacts on the core stability if controlling for the mentioned explanatory variables. The average stability of the cores in the Natural sciences and mathematics, Biotechnological sciences and Social sciences is higher than the average core stability of the Humanities, but not statistically significant. On the other hand, the average value of the stability of the cores in Engineering sciences and technologies and Medical sciences is statistically significantly higher than in the Humanities.

The goal of this study was to measure the extent of the persistence of co-authorship collaboration of researchers and analyse the factors that explain this persistence. The study was based on administrative data on all active researchers available in the Slovenian research system. Combining these data with data obtained from a survey for at least some selected scientific disciplines would allow a deeper insight into the analysed phenomenon.

**Acknowledgements** The authors appreciate the comments given by the anonymous reviewers. This research has been supported by the Slovenian Research Agency (J5-5537).

## References

- Abbasi A, Altmann J, Hossain L (2011) Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics* 5(4):594–607
- Abbasi A, Hossain L, Leydesdorff L (2012) Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics* 6(3):403 – 412
- Adams JD, Black GC, Clemmons JR, Stephan PE (2005) Scientific teams and institutional collaborations: Evidence from u.s. universities, 1981–1999. *Research Policy* 34(3):259 – 285
- Albatineh AN, Niewiadomska-Bugaj M, Mihalko D (2006) On similarity indices and correction for chance agreement. *Journal of Classification* 23(2):301 – 313
- Batagelj V, Doreian P, Ferligoj A (1992a) An optimizational approach to regular equivalence. *Social Networks* 14(1):121–135
- Batagelj V, Ferligoj A, Doreian P (1992b) Direct and indirect methods for structural equivalence. *Social networks* 14(1):63–90
- Batagelj V, Ferligoj A, Doreian P (1998) Fitting pre-specified blockmodels. In: Hayashi C, Yajima K, Bock HH, Ohsumi N, Tanaka Y, Baba Y (eds) *Data Science, Classification, and Related Methods, Studies in Classification, Data Analysis, and Knowledge Organization*, Springer Japan, pp 199–206
- Beaver D (2004) Does collaborative research have greater epistemic authority?. *Scientometrics* 60(3):399 – 408
- Beaver D, Rosen R (1978) Studies in scientific collaboration. *Scientometrics* 1(1):65–84
- Borrett SR, Moody J, Edelman A (2014) The rise of network ecology: Maps of the topic diversity and scientific collaboration. *Ecological Modelling* 293:111–127
- Burt RS, Minor MJ, Alba RD (1983) *Applied network analysis: A methodological introduction*. Sage Publications Beverly Hills, CA
- Chinchilla-Rodríguez Z, Ferligoj A, Miguel S, Kronegger L, de Moya-Anegón F (2012) Blockmodeling of co-authorship networks in library and information science in argentina: a case study. *Scientometrics* 93(3):699–717
- Cugmas M, Ferligoj A (2015) On comparing partitions. Submitted
- De Haan J, Leeuw F, Remery C (1994) Accumulation of advantage and disadvantage in research groups. *Scientometrics* 29(2):239–251
- Doreian P, Batagelj V, Ferligoj A (1994) Partitioning networks based on generalized concepts of equivalence. *Journal of Mathematical Sociology* 19(1):1 – 27
- Ferligoj A, Kronegger L (2009) Clustering of attribute and/or relational data. *Advances in Methodology & Statistics* 6(2):135–153
- Ferligoj A, Kronegger L, Mali F, Snijders T, Doreian P (2015) Scientific collaboration dynamics in a national scientific system. *Scientometrics* 104(3):985 – 1012
- Frenken K, Hözl W, de Vor F (2005) The citation impact of research collaborations: the case of European biotechnology and applied microbiology (1988–2002). *Journal of Engineering and Technology Management* 22(1–2):9–30
- Friedkin NE (1981) The development of structure in random networks: an analysis of the effects of increasing network density on five measures of structure. *Social Networks* 3(1):41–52
- Garg K, Dwivedi S (2014) Pattern of collaboration in the discipline of japanese encephalitis. *DESIDOC Journal of Library & Information Technology* 34(3)

- Glänzel W (2002) Coauthorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies. *Library Trends* 50(3):461
- Groboljšek B, Ferligoj A, Mali F, Kronegger L, Iglíč H (2014) The role and significance of scientific collaboration for the new emerging sciences: the case of Slovenia. *Teorija in Praksa* 51(5):866 – 885
- Hollis A (2001) Co-authorship and the output of academic economists. *Labour Economics* 8(4):503–530
- Howells J, James A, Malik K (2003) The sourcing of technological knowledge: distributed innovation processes and dynamic change. *R&D Management* 33(4):395–409
- Hu Z, Chen C, Liu Z (2014) How are collaboration and productivity correlated at various career stages of scientists?. *Scientometrics* 101(2):1553–1564
- Katz JS, Martin BR (1997) What is research collaboration? *Research policy* 26(1):1–18
- Kronegger L, Ferligoj A, Doreian P (2011) On the dynamics of national scientific systems. *Quality & Quantity* 45(5):989–1015
- Kronegger L, Mali F, Ferligoj A, Doreian P (2015) Classifying scientific disciplines in Slovenia: A study of the evolution of collaboration structures. *Journal of the Association for Information Science and Technology* 66(2):321–339
- Kyvik S (2003) Changing trends in publishing behaviour among university faculty, 1980-2000. *Scientometrics* 58(1):35–48
- Lambiotte R, Panzarasa P (2009) Communities, knowledge creation, and information diffusion. *Journal of Informetrics* 3(3):180–190
- Laudel G (1999) Interdisziplinäre Forschungsk Kooperation: Erfolgsbedingungen der Institution 'Sonderforschungsbereich' / Interdisciplinary research cooperation : success conditions of the 'special research field' institution. Ed. Sigma
- Laudel G (2002) What do we measure by co-authorships? *Research Evaluation* 11(1):3–15
- Lee S, Bozeman B (2005) The impact of research collaboration on scientific productivity. *Social studies of science* 35(5):673–702
- Liu X, Bollen J, Nelson ML, Van de Sompel H (2005) Co-authorship networks in the digital library research community. *Information processing & management* 41(6):1462–1480
- Lorrain F, White HC (1971) Structural equivalence of individuals in social networks. *The Journal of mathematical sociology* 1(1):49–80
- Lundberg J, Tomson G, Lundkvist I, Skår J, Brommels M (2006) Collaboration uncovered: Exploring the adequacy of measuring university-industry collaboration through co-authorship and funding. *Scientometrics* 69(3):575–589
- 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* 1(2)
- Mali F, Kronegger L, Doreian P, Ferligoj A (2012) Dynamic scientific co-authorship networks. In: *Models of Science Dynamics*, Springer, pp 195–232
- Melin G (2000) Pragmatism and self-organization: Research collaboration on the individual level. *Research policy* 29(1):31–40
- 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 (2004) Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States* (14):5200
- Nooy Wd, Batagelj V, Mrvar A (2011) *Exploratory social network analysis with Pajek. Structural analysis in the social sciences*: 34, Cambridge University Press
- Price DJdS (1963) *Little science, big science*. Columbia University Press

- White DR, Reitz KP (1983) Graph and semigroup homomorphisms on networks of relations. *Social Networks* 5(2):193–234
- Yoshikane F, Nozawa T, Tsuji K (2006) Comparative analysis of co-authorship networks considering authors' roles in collaboration: Differences between the theoretical and application areas. *Scientometrics* 68(3):643–655

This is manuscript.