Original scientific paper

"SCIENTIFIC COLLABORATIONS" AT THE LEVEL OF COUNTRIES – A CASE STUDY: A "HARD" PHYSICAL GEOGRAPHY AND A "SOFT" ECONOMICS SCIENCE FIELD

Zsófia Viktória VIDA[a]

[a] Department of Science Policy and Scientometrics, Library and Information Centre of the Hungarian Academy of Sciences, vida.zsofia@konyvtar.mta.hu

Abstract

Scientific collaboration is a widely studied science field in the recent decades. For measuring, it a tighter approach co-authorship is frequently used. Co-authorship base level is the level of articles and the co-authorship networks are based on articles and authors. In this study a rather different level was examined: the level of journals. We used the approach of collaboration analysis regarding the relation between countries and journals. Two main groups of methods were used: statistical analysis and networks analysis. The goal was to explore patterns between journals and countries. We investigated the relations between journal ranking and affiliated countries. We were interested if there are any differences between a "hard" and a "soft" science field. Both of our methodological approaches showed similar answers to research questions. The differences were minimal between "hard" and "soft" science fields. The journals with the highest and the lowest journal ranking had the highest inequalities in the distribution of countries.

Keywords: science network, scientific collaboration, journal affiliation, journal ranking

INTRODUCTION

Scientific collaboration is a widely studied science field in the recent decades (Acosta et al., 2011; Boschma, 2005). Several types of proximities help in building up science networks and the intensity of collaborations. Boschma (2005) distinguishes five types of proximities (geographical, cognitive, institutional, organizational, and social). Narin, Stevens and Whitlow (1991) also emphasize the linguistic and cultural aspects in scientific relations. In this study the focus is on geographical and social proximity. This article is not about a real type of collaboration: only the aspects of scientific collaboration are used in this study.

First of all, we have to give the definition of scientific collaboration. The dictionary definition of collaboration is 'two or more individuals working together to reach a common goal' (http://www.businessdictionary.com/definition/collaboration.html). Therefore, scientific collaboration could mean working together, co-operating to achieve a common

research goal: producing a new scientific knowledge. (Katz, Martin, 1997). This definition is a quite wide interpretation of collaboration. Measuring collaboration is the most obvious issue if we focus only on co-authorship. Co-authorship is not equal to research collaboration. Scientists can collaborate for example by working together without writing a common scientific article. On the other hand, this type of collaboration can be measured in a difficult way. Co-authorship is a tighter interpretation of research collaboration. In the following this tighter interpretation will be used for scientific collaborations.

Research collaboration can be studied through the relations of authors and the relations of authors' affiliation data as well. Authors' affiliation data are the information, which contains the address of the author of an article. Authors' affiliation data can be distinguished by several scales of the contributor actors into the following types: departments, institutions, settlements, regions, and countries. The international collaboration can be measured by these data (Braun et al., 1992, Luukkonen et al., 1993, Schubert, Braun, 1990). Thus, co-authorship base level is the level of articles and the co-authorship networks are based on articles and authors. Co-authorship networks can be examined by different scales where the network shows the relation between articles and different scales of authors' affiliation data e.g. articles and institutions, articles and regions and so on. Examining co-authorship networks at a more and more aggregate level can be analyzed from authors to countries.

In this study the level of articles will not be analyzed. A rather different level was examined: the level of journals. The original goal was a network study, based on how different countries contribute to journal content in this methodological framework. Are there any patterns, systematic structure formations? At the usual level of authors and articles, it seems to be a collaboration study. However, at the level of journals and countries it is completely different. That is why scientific collaborations are in quotation marks in the title. This is a special pseudo-collaboration and not a real type of collaboration.

In our case the base components are journals and not articles, therefore the analysis focuses on journals and countries. The countries derive from journals, which contain articles and the authors' affiliation data. Also in these types of network the nodes can be countries but the links are the common journals (and not the common articles).

There is a relation between research collaboration and papers' citation impact. The papers, which have more co-authors may have higher papers' citation impact (Frenken et al., 2005, Goldfinch et al., 2003). In this study the base components are journals, therefore

journal ranking was involved in the analyses. Journal ranking as the measure of quality of a journal is an important factor. Nowadays journal evaluation and ranking are often used in different situations. Citations of a journal show its impact. The absolute number of citations depends on many factors such as time or discipline. The most well-known size-independent measure is the journal impact factor of Garfield (1972). The impact factor is the division of the number of citations during a given time period and the number of articles during a time period. The original impact factor used two year time periods but five year time periods are also used. Impact factor and other journal citation metrics are presented every year in Journal Citation Rank (JCR) by Thomson Reuters (Haustein, 2012). The value of the impact factor depends on science fields because of the different citation habits. To compare the impact factor measure between different science areas, field normalization is required (Sen, 1992, Pudovkin, Garfield, 2004). In this study the Normalized Journal Position (NJP) was used. This is the science field normalized impact factor by Thomson Reuters.

In the following sections, first of all the questions and objectives of the study will be presented. After that, the data of the study and the used methods will be stated. Finally, the results and conclusions will be presented.

OBJECTIVES AND METHODS

In this study the relation between journals and countries will be investigated. The countries are contained in the authors' affiliation data in journal articles. Usually this phenomenon is examined at the level of articles when it can be described with collaboration networks (Hoekman et al., 2010, Hou et al., 2008). In this case we do not talk about a collaboration network. When the relation between journals and countries is examined it can be described also with a network that is similar to collaboration networks. It can be called a pseudo-collaboration network. The study uses the approach of research collaboration to analyze the patterns of this pseudo-collaboration.

The goal of the study is to explore patterns in these relations especially in Social Science and Humanities (SSH) fields. The study focuses on one chosen SSH field, namely Economics. This science area was compared to Physical Geography, which is a natural science. Therefore, across these two science fields the SSH and natural science fields could be compared. Are there any characteristic patterns? In this relation are there any differences between "hard" and "soft" science fields?

Every journal has an editorial board, which chooses the articles for publishing. The chosen articles' affiliation data contain the authors' country data as well. In this study, as mentioned above, we focus on the country affiliation data.

Our research questions are the following: Are there any patterns at the level of journals in the countries' data? What kind of patterns are there at this level? Are there any relations between the journal ranking and affiliated countries? Are the countries, which investigate more R&D or GDP more successful during the peer review procedure?

Development is a multi-dimensional and multi-indicator phenomenon (Nemes Nagy, 2009). Dozens of indicators and methods exist for measuring development. The most generally used indicator to measure development is Gross Domestic Product (GDP) or GDP per capita. Defining development is out of the problematics of the present study. We are interested if there is any relation between the pseudo-collaboration network and development. The countries, which have a central position in the network, are the countries, which have high GDP per capita as well. This question is analyzed by network analysis.

Data

This study explored the patterns between journals and countries. We used data coming from Web of Science (WoS). WoS is an online research platform powered by Thomson Reuters. WoS provides the most known scientific citation databases. It includes seven citation databases: Science Citation Index Expanded, Social Science Citation Index, Art & Humanities Citation Index, Index Chemicus, Current Chemical Reactions, Book Citation Index and Conference Proceedings Citation Index. The content of the database is current and retrospective in the fields of Science, Social Science, Arts and Humanities. It contains more than 120 000 journals worldwide with the highest impact level. It contains the main data of each publication, namely title, author(s), keywords, abstract, Web of Science Categories (WCs), authors' affiliation, publication year, journal title, citation indexes and so on (wokinfo.com). The content of the database is overrepresented by English language journals.

The examined two fields of science are a "soft" one, Economics and a "hard" one, Physical Geography. To determine the science fields WoS categorization system was used where more than 250 Web of Science Categories (WCs) are distinguished in Science, Social Science and Arts & Humanities (incites.isiknowledge.com). In the study "WC"

abbreviation is used for WoS Categories. WC is the lowest level of aggregation of science fields (Leydesdorff, Carley, Rafols, 2013).

The study examined journals from the chosen WCs. Every journal is classified to one or more WC. Based on common WCs a co-occurrence matrix was created. Not only the Economics and Physical Geography WCs were used but also those WCs which are in strong relation with Economics or Physical Geography. Thus, more journals could be included to the investigation. Therefore, in choosing the WCs hierarchical clustering was used on the distance matrix to investigate more journals in the study. The distance matrix derived from the journal and WC co-occurrence matrix. We clustered the WCs. The used cut level was 80% with Physical Geography and 90% with Economics. From 50 to 95% cut levels were examined in both science fields. The chosen cut levels were 80 and 90% because these levels contain not only Economics or Physical Geography and nor too much WCs. The cluster which contained Economics and Physical Geography was chosen. Thus the chosen WCs are the following:

Economics:

- Agricultural Economics & Policy;

- Business, Finance;

- Economics;

Physical Geography:

- Geography, Physical;

- Geosciences, Multidisciplinary;

- Imaging Science & Photographic Technology;

- Remote Sensing;

- Engineering, Geological

Journals from the chosen WCs were examined. Thus, there were 238 journals from the fields of Physical Geography group, hereinafter Physical Geography and 384 from the Economics group, hereinafter Economics.

In the study the focus was on the distribution of countries based on author affiliations for every article between 2010 and 2014 for each chosen journal. Fig. 1 shows an example of country distributions in one journal. This picture is typical: some countries get a lot of country affiliations and many countries get a few ones.

Figure 1 Country distribution of one journal
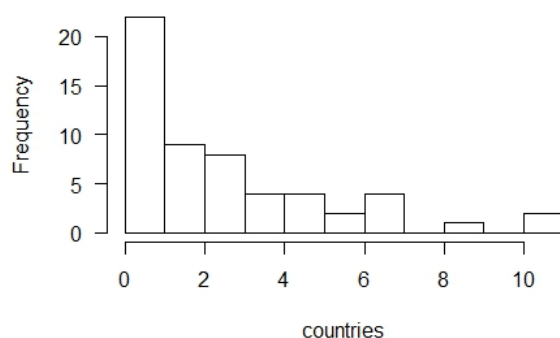
Methods

The examined patterns between journals and countries were explored by two main groups of methods: statistical analysis and network analysis. First of all, the entire picture was analyzed to find patterns. The problem was approached from two ways: by statistical methods and by network analyses. Within statistical methods to investigate patterns between journals and countries the used methods were GINI coefficient and hierarchical clustering. To explore the country distribution inequalities, the used method was GINI coefficient. For each journal the GINI coefficient was calculated to the country distribution. To find different journal groups the used method was hierarchical clustering. Hierarchical cluster analysis was based on journal and country distance matrix.

To find more detailed relations among data, journal ranking was involved in the analyses. To answer one of the research questions – 'Are there any relations between the journal ranking and affiliated countries?' – the journals were divided in four groups by Normalized Journal Position (NJP). This indicator lets us to compare different science fields. This is the science field normalized impact factor by Thomson Reuters. The used method was comparing NJPs to NJPs' average, therefore the journals were sorted in two categories and after that the procedure was repeated. The relation between the GINI coefficient and journal rank was analyzed. The relation between the GINI coefficient and country clusters were analyzed as well.

To answer the research question – 'Are the countries which investigate more R&D or GDP more successful during the peer review procedure?' – we analyze the relation between the GDP data and the network of countries (pseudo-collaboration network). We

used network visualization to see if the countries in central position of the network are the same countries with higher GDP per capita.

To examine the data jointly the used method was network analysis. The questions were also examined from a different point of view: network analysis. To this method first of all it is necessary to determine the network's nodes and edges (Csermely, 2005, Newman 2003). The examined phenomenon can be directly described with a "two-mode network". In this type of network, the relation between journals was established through the common countries. Fig. 2 shows a model of this type of networks. J means journals and C means countries. The relation between two journals is across countries. In two-mode networks the nodes belong to sets. There are no direct relations between the nodes which belong to one set (Barabási, 2014, Wasserman, Iacobucci, 1991). Now there are no relations between journals and there are no relations between countries. The nodes which belong to country type created the relation between journals.

Figure 2 Two-mode network model
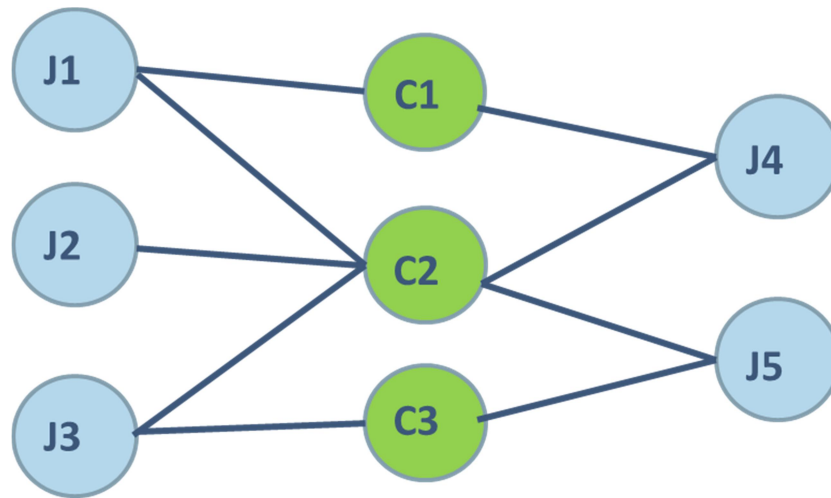
However, it can be also described with a normal (one-mode) network, where the nodes are the countries and the edges are relations between two countries due to a journal. To this we converted the two-mode network to one-mode (Everett, Borgatti, 2013). A link/edge exists between two countries when there is at least one journal which country affiliation distribution contains the two given countries. This network was weighted and undirected. Stronger relation between two countries means that these countries were chosen by several journals. The stronger relation was between two countries, the more journal country affiliation distribution contained the two given countries.

For calculations several packages of R software were used (Csardi, Nepusz, 2006, Meyer, Buchta, 2015, Langfelder, Zhang, 2014, R Core Team, 2015). For network visualization Gephi software (Bastian et al., 2009) was used.

RESULTS

To answer the former questions in this section our results will be presented. To detect patterns, the focus was on two aspects: the relation between the journal ranking and affiliated countries and the relation between GDP and countries' position.

Our first question was the following: is there any relation between the journals' country distribution and journals' quality? To describe the inequalities of country distribution the calculated indicator was GINI coefficient for each journal. The lower GINI coefficient indicates more balanced country structure. Every examined journal by Normalized Journal Position (NJP) was categorized to four groups (1 to 4), as mentioned above. The best rank was 1. Fig. 3 and 4 show the relation between GINI coefficient and journal average rank.

Figure 3 Boxplot of GINI and Journal average rank on the field of Economics
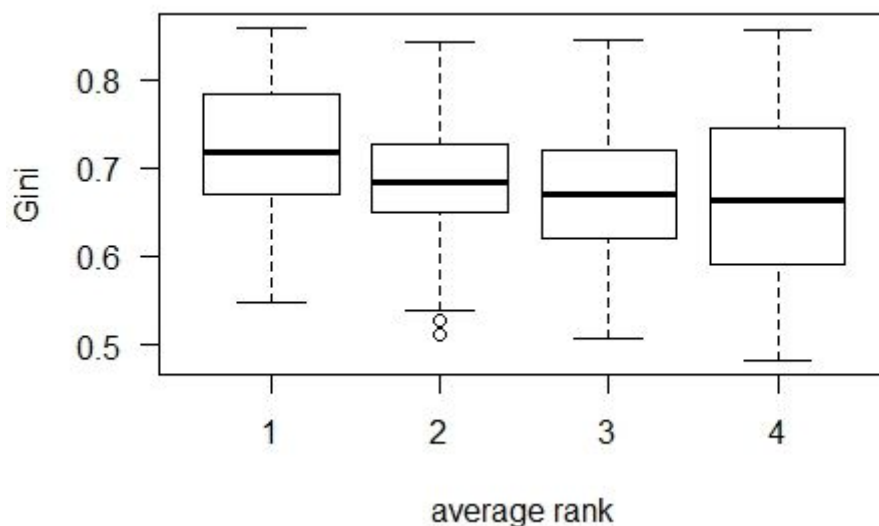
Figure 4 Boxplot of GINI and Journal average rank on the field of Physical Geography
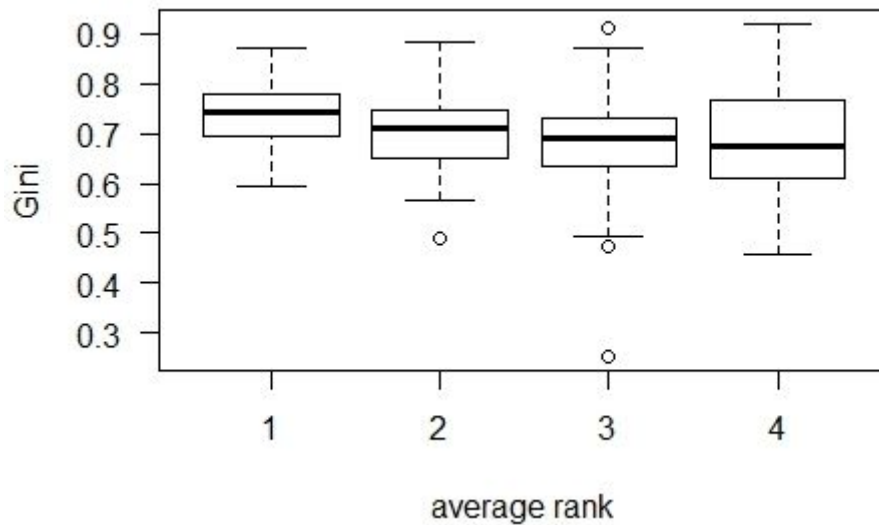
The central tendency decreased a bit from '1' average journal rank category to '4' in both cases. The GINI coefficients had the highest value in journals, which belong to the first ranking group. It means that the "better" journals have more inequalities among the chosen countries. In this case "chosen countries" mean that the articles were selected by journal editorial board and this choice determines countries by authors' affiliations. The deviation was the highest in the fourth group and also high in the first group. Therefore the best and the weakest journals had the most different GINI indexes.

Fig. 3 and 4 consider directly journals. In the next step the analysis considers the clusters of journals. To detect journal groups, the used method was hierarchical clustering. Hierarchical cluster analysis was used based on journal and country distance matrix. With the obtained journal clusters the relation between clusters and journal ranking was investigated. Between journals and countries, a matrix was determined. Tab. 1 shows a part of the matrix. The cells contain the amount of distribution of the countries.

Table 1 A part of the journal-country matrix

country	journal ISSN number				
	05701864	07173830	07185286	0734306X	07350015
UKRAINE	0.000	0.000	0.000	0.000	0.000
UNITED KINGDOM	0.000	0.000	0.000	0.000	0.000
URUGUAY	0.000	1.471	4.167	0.000	0.000
USA	33.247	14.706	16.667	67.424	60.656

From this matrix a similarity matrix was calculated. The used method was Cosine similarity (Nguyen, Bai, 2010). Cosine similarity gives the coherence between two vectors.

In co-citation analysis Salton's Cosine similarity is often used. Fig. 5 shows the formula of Salton's Cosine similarity in the case of citation analysis (Hamers et al., 1989) where coc(i,j) is the number of co-citation between i and j; cit(i) is the number of citation of i and cit(j) is the number of citations of j. The value of Ss(i,j) is between 0 and 1. Salton's cosine formula is undefined when cit(i)=0 or cit(j)=0. The similarity is complete and the Ss(i,j) value is 1, when cit(i) and cit(j) and coc(i,j) are equal. There is no similarity and the Ss(i,j) value is 0 when there is no co-citation (coc(i,j)=0). In case of Salton's Cosine similarity, the number of common citations is analyzed between i and j articles. In our case the counter is the number of common countries between i and j journal. cit(i) is the number of countries in i journal and cit(j) is the number of countries of j journal.

Figure 5 Salton's cosine formula

$$S_S(i,j) = \frac{coc(i,j)}{(\text{cit}(i) \cdot \text{cit}(j))^{1/2}}.$$

Source: Hamers et al., 1989 p. 315

Tab. 2 shows a part of this similarity matrix. The higher cell content means that between two journals there are more common chosen countries. The values at the cells can vary between 0 and 1 as mentioned above. The distance matrix was determined from the similarity matrix. (d=1-s, where d is distance matrix and s is similarity matrix). To classify the journals hierarchical cluster analysis was used. Different methods/types of hierarchical cluster analysis (ward, single, complete, average, mcquitty, median, centroid) were compared. The average method turned out to be the best one because the distance between the distance matrix and dendrogram was the least in this case.
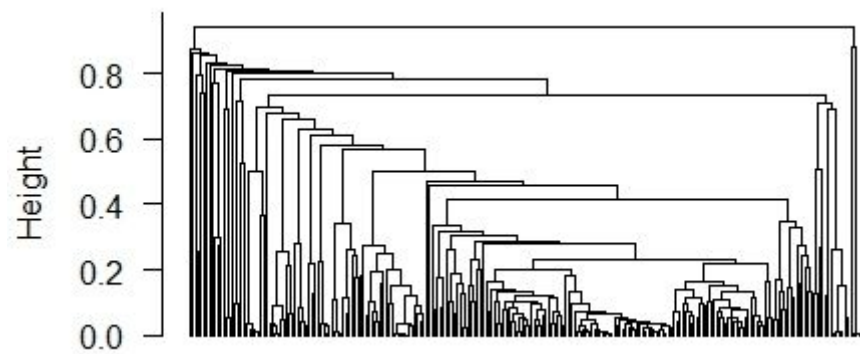
Table 2 A part of the journal-journal similarity matrix

	0002-9599	0003-813X	0004-0843	0008-3674	0008-4077
0002-9599	0	0,3621007	0,4860055	0,2547559	0,5986628
0003-813X	0,3621007	0	0,6514876	0,4356771	0,6938004
0004-0843	0,4860055	0,6514876	0	0,2136618	0,0254593
0008-3674	0,2547559	0,4356771	0,2136618	0	0,2048588
0008-4077	0,5986628	0,6938004	0,0254593	0,2048588	0

Fig. 6 and 7 show the dendrograms of the two chosen science fields. Two aspects were applied to create the groups of journals. The first approach was to cut the tree at a fix points, the second one was to use dynamic cut tree procedure.

In the first case the selected cutting level was 0.7. Cutting level was examined between 0.5 and 0.9. 0.7 cutting level seemed to be the best choice. Fig. 8 and 9 present the clusters with their sizes. Cluster structures and cluster sizes were similar (power-law like distribution) in the two science fields. Three groups were created from the clusters. Thus, the clusters with similar sizes were analyzed together. The first one contained the most of the journals, it was the biggest cluster. The second group contained several clusters, each of them had some journals. In the third group were those journals which belonged to separate clusters.

Figure 6 Dendrogram of Economics journals
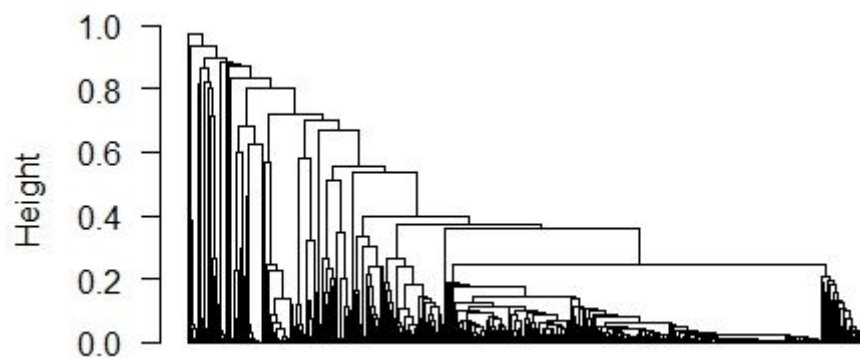
Figure 7 Dendrogram of Physical Geography journals

Figure 8 Clusters of Economics journals

Figure 9 Clusters of Physical Geography journals

The second approach to create groups of journals is the dynamic cut tree procedure. The dynamic cut tree procedure regards the structure of the dendrogram. It does not use a fix level to cut the tree. (Langfelder et al., 2008, Langfelder, Zhang, 2014). Using this procedure, the given clusters were the following (Tab. 3 and Tab. 4):

Table 3 Clusters from Dynamic Cut Tree algorithm in Economics

| Economics	
Cluster id	Cluster size (number of journals)
1	102
2	96
3	56
4	49
5	47
6	33
7	1

Table 4 Clusters from Dynamic Cut Tree algorithm in Physical Geography

| Physical Geography	
Cluster id	Cluster size (number of journals)
1	88
2	78
3	43
4	23
5	6

In the next step the analysis considers cluster groups of journals. First the groups of clusters were analyzed, later the clusters, which derived from dynamic cluster procedure.

Fig. 10 and 11 present the relation between GINI coefficient and the groups of clusters. There were some differences between the two science fields. The field of Economics showed higher deviation in each group. The third group, which contained the separated clusters showed the highest deviation in the field of Economics, and the lowest in the field of Physical Geography. The journals in the third group which belonged to separate clusters in the field of Physical Geography were similar to each other. The same statement cannot be told about the field of Economics. There were not huge differences in the value of GINI among the cluster groups. Therefore, the distribution of affiliated countries in each cluster was similar. In this aspect there were not big inequalities.

Figure 10 Boxplot between clusters groups and GINI coefficient in Economics journals
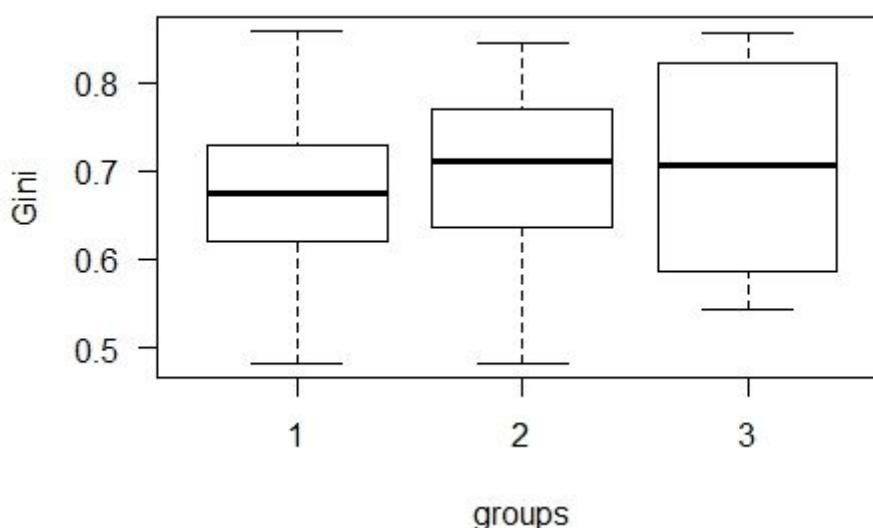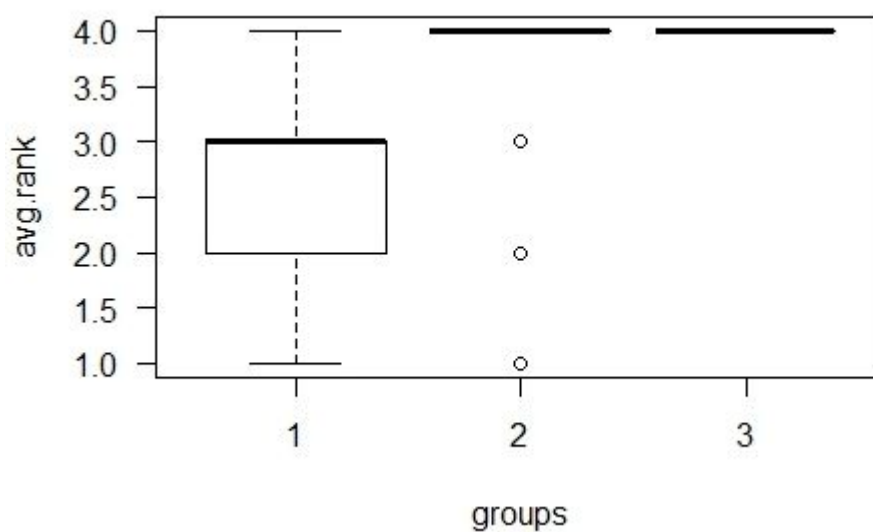
Figure 11 Boxplot between clusters groups and GINI coefficient in Physical Geography journals
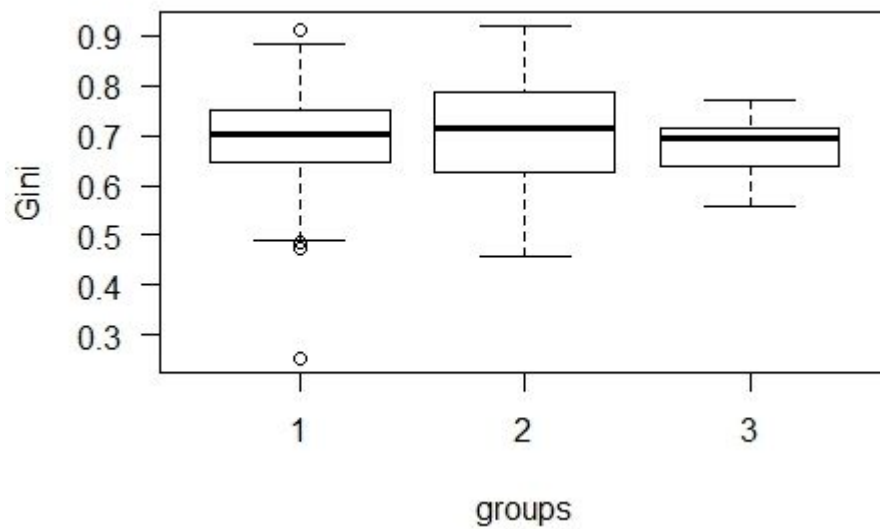

Contrarily, the journal ranks showed differences between the cluster groups. Fig. 12 and 13 present the boxplot between cluster groups and journals' average rank in the two chosen science fields. Second and third cluster groups contained "weaker" journals in both science fields. The first cluster group, which contained most journals shows the highest deviation for average journal rank. Therefore, the relation between journal ranks and GINI coefficient did not have a big bias between the cluster groups.

Figure 12 Boxplot between cluster groups and journal average rank in Economics journals

Figure 13 Boxplot between cluster groups and journal average rank in Physical Geography journals
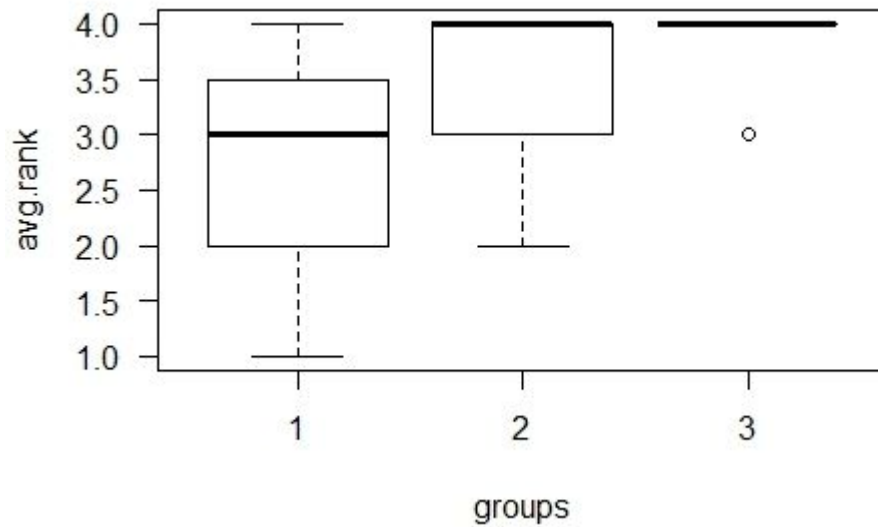
The clusters of dynamic cut tree procedure were presented above in Tab. 3 and 4. The content of each cluster was summarized. In each cluster the distribution of country affiliation of journals was summarized. The distribution had a very long tail, therefore the data were cut at the upper quartile (Q3=178,5 in Economics and 152,5 in Physical Geography). The clusters' country distribution (Fig. 14, 15) showed the same picture as the journals (Fig. 1).

Figure 14 Country distribution in the journals of cluster 1 in Economics
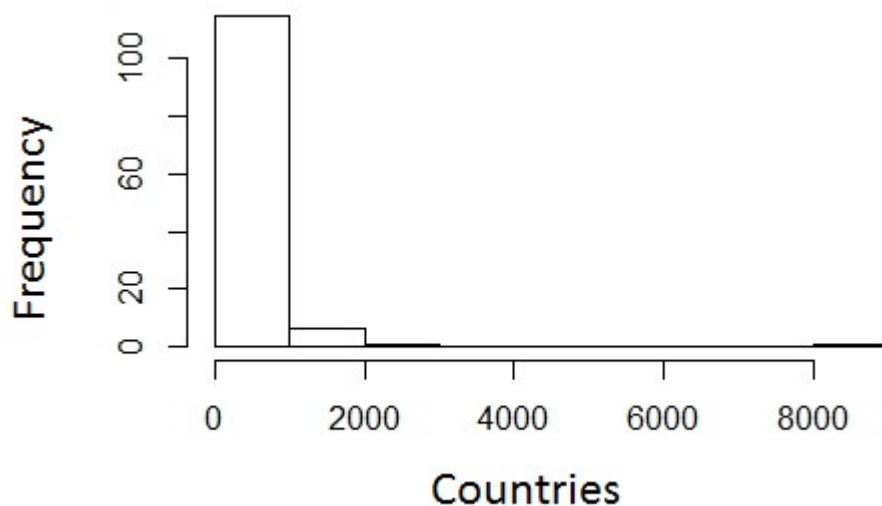
Figure 15 Country distribution in the journals of cluster 1 in Physical Geography
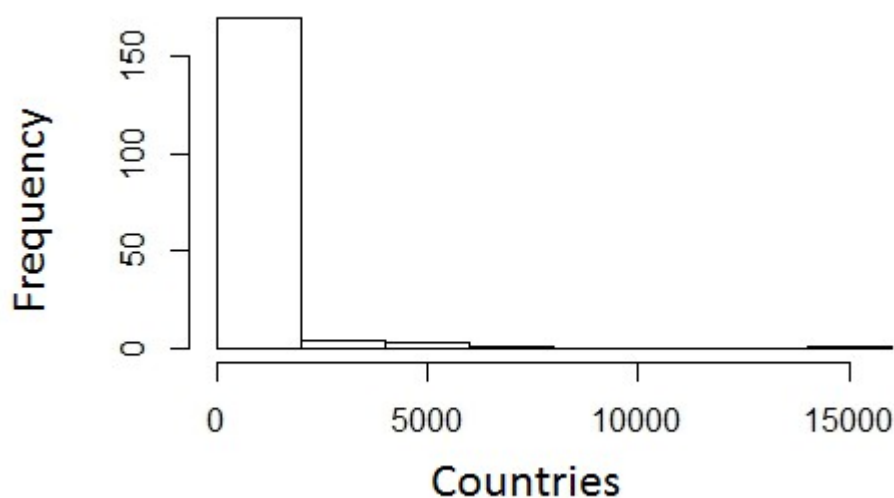
Tab. 5 and 6 show each clusters' first 20 countries by the ranking of clusters' country frequency. The order of countries did not show big differences at the first three places. There was a bit more variability in the order of countries at Physical Geography. China had better place in ranking at Physical Geography than Economics. In developing countries "hard" Science spread easier than Social Sciences.

Table 5 Country frequency rank clusters in Economics

Ország	1. cluster	2. cluster	3. cluster	4. cluster	5. cluster	6. cluster
USA	1	1	1	1	1	1
ENGLAND	2	2	2	2	2	2
GERMANY	3	3	3	3	3	3
AUSTRALIA	4	6	7	6	8	4
SPAIN	5	8	8	7	5	11
ITALY	6	7	9	8	10	9
FRANCE	7	5	5	10	4	10
CANADA	8	10	4	4	6	8
PEOPLES R CHINA	9	4	10	9	7	6
NETHERLANDS	10	9	6	5	9	5
CZECH REPUBLIC	11	37	34	27	22	26
JAPAN	12	11	18	16	12	7
SWITZERLAND	13	17	11	11	11	13
BELGIUM	14	16	12	14	13	15
TAIWAN	15	12	15	15	23	12
SWEDEN	16	13	16	13	14	17
SOUTH KOREA	17	15	14	18	16	14
DENMARK	18	19	20	19	18	21
NORWAY	19	20	35	17	19	22
SLOVAKIA	20	37	35	27	34	26

Table 6 Country frequency rank clusters in Physical Geography

Ország	1. cluster	2. cluster	3. cluster	4. cluster	5. cluster
USA	1	1	1	2	1
PEOPLES R CHINA	2	2	2	1	2
GERMANY	3	4	3	4	5
ENGLAND	4	6	6	3	4
FRANCE	5	3	5	8	3
CANADA	6	9	7	6	6
AUSTRALIA	7	12	8	5	7
ITALY	8	8	4	9	10
SPAIN	9	11	9	12	8
JAPAN	10	10	10	7	16
NETHERLANDS	11	13	12	13	15
SWITZERLAND	12	14	13	14	11
INDIA	13	7	11	10	23
SOUTH KOREA	14	26	14	16	31
NORWAY	15	17	18	11	26
SWEDEN	16	19	19	20	14
TAIWAN	17	21	15	18	32
RUSSIA	18	5	17	27	32
SCOTLAND	19	25	27	19	21
BRAZIL	20	16	22	17	17

The relation between journals and country affiliation data were examined from a different aspect as well. Network aspect helps us to discover the relations between certain phenomena. Tab. 7 presents the main network attributes. The network was very dense (72% and 82%). Network density can be described with a number, which refers to the whole network. It is the ratio of the number of relations in the network and of the number of the potential relations in the network (n*(n-1)/2), where n is the number of nodes (Csermely, 2005, Newman, 2003).

Table 7 The main data of networks

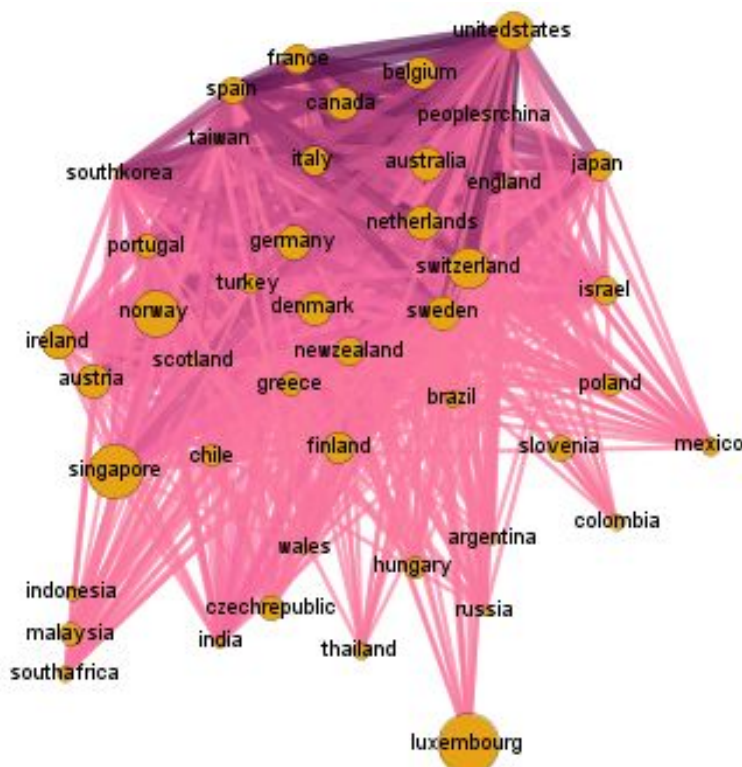|	Economics	Physical Geography
number of nodes	171	194
number of edges	10 470	15 356
network density (%)	72	82

Tab. 8 presents the main statistic parameters of edge weights. The edge weights showed a power-law distribution. The distribution had a very long tail. The edges of networks were filtered. The edges below the upper quartile were deleted from the network. Thus the number of edges was reduced to approximately a fifth (2118 in Economics and 3722 in Physical Geography). The number of nodes, which had at least 1 degree (the non-isolated nodes) were 77 in Economics and 106 in Physical Geography. The degree of a node is the number of its links (Csermely, 2005, Newman, 2003).

Table 8 The statistic of edge weight of networks

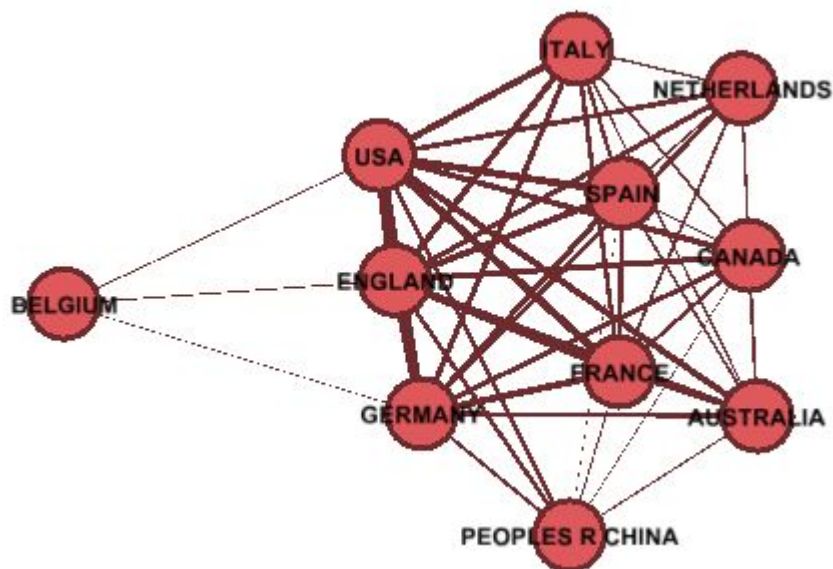|	Economics	Physical Geography
minimum	1	1
lower quartile	1	2
median	4	7
upper quartile	15	20
maximum	375	231

Fig. 16 and 17 show the visualization of networks. The size of the nodes were proportionate to GDP per capita and the colors of edges were proportionate to edge weight. The darker was the color, the stronger was the relation between two countries. The centrum countries had stronger links between each other than others. The countries on the periphery (on the bottom of the figures) were connected more to the centrums than to each other. A centrum-periphery structure was outlined. The GDP per capita was not so relevant in the relations. The other relevant aspect was the population of the country. The size was the strongest feature in the relations between countries.

Figure 16 The network of countries at the fields of Economics

Figure 17 The network of countries at the fields of Physical Geography
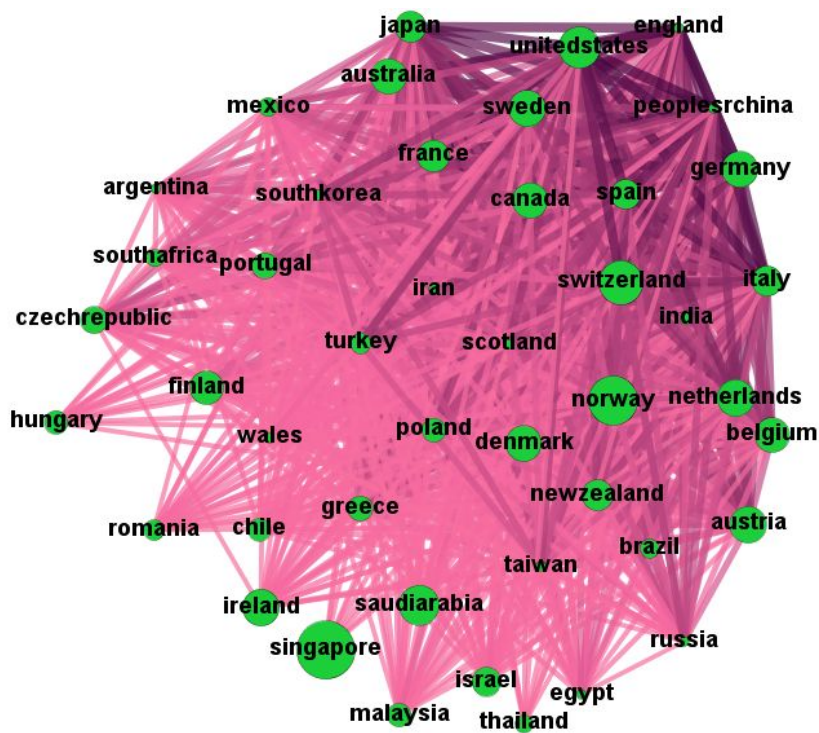
Figure 18 The „top" network of countries at the fields of Economics

Figure 19 The „top" network of countries at the fields of Physical Geography
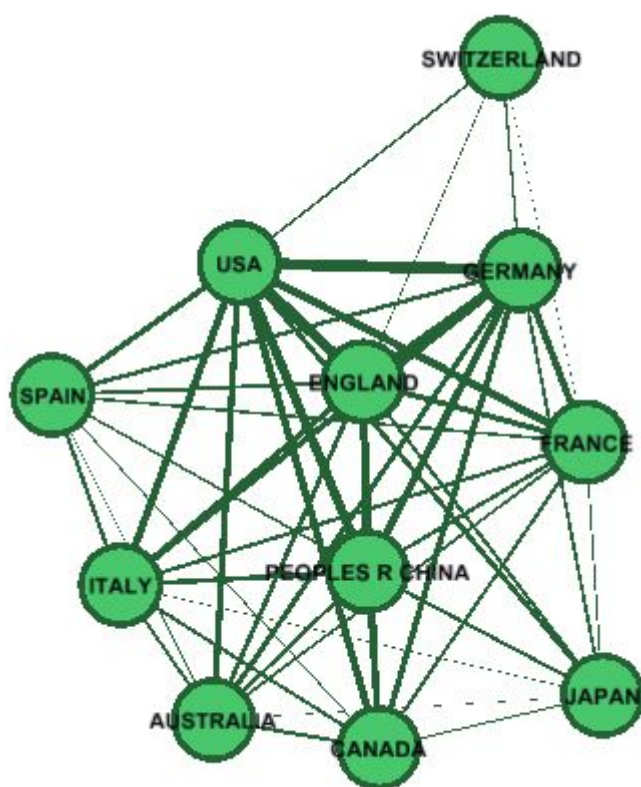

Fig. 18 and 19 show the top 1 percent of the edges' weight. Therefore, we can see the most frequently chosen countries. These countries are the centrum countries. China's presence could be explained with its population size.

DISCUSSION AND CONCLUSION

In this study the relation between journals and countries by the authors' affiliation data was examined. This study is rather a methodological one. The goal was to explore the patterns and systematic formations in these relations, to observe whether there are any differences between "hard" and "soft" science fields. The observation level was an uncommon one, the level of journals. Usually the relation between authors' affiliation was analyzed at the level of articles by using collaboration networks. At the level of journals this relation is completely different but it can be analyzed in a similar way with networks as well. The approach of collaboration analysis was used to these pseudo-collaborations.

All methods served to answer whether there are any patterns at the level of journals on the countries' data. Two main analysis aspects were used in this study to answer the research questions: statistical aspects and network aspects. Both methods gave similar answers. There were similar patterns at the level of journals than the usually analyzed level of articles.

Between the two science fields the differences were minimal. The "hard" and "soft" science showed the same tendencies in the relation between countries and journals.

To detect patterns two aspects were deeply analyzed. These are the following: the relations between the journal ranking and the affiliated countries; and the relation between countries' network positions and countries' GDP per capita.

At the aspect of relation between the distribution of countries (GINI coefficient) and journal average rank the journals which had higher ranking had more inequalities between the chosen countries. The best and the weakest journals had the most different GINI indexes.

Tab. 5 and 6 show the most frequently appearing countries. At the order of the countries "hard" science showed bigger variety. India also appears among developing countries in the rank of countries. At the fields of hard science developing countries can have a bigger chance to appear in science mainstream.

The results of network analyses showed the same patterns. In Physical Geography the presence of China is surprising, but the size of its population can explain it. A centrum-periphery structure was outlined in both networks.

From the point of view of the relation between countries' network positions and countries' GDP per capita, those countries which had higher GDP per capita show higher frequency. There were some except those, which have high values because the population is low, for example Singapore or Luxembourg. Beside GDP the population of countries was also a main factor in the relations between countries.

In the network, which presented the part of the network containing the top 1% of edges' weight there were the centrum countries. These countries in Fig. 18 and 19 are mostly the same than the top 10 countries in Tab. 5 and 6. The main Anglo-Saxon countries – USA, England, Australia, Canada – were in the top 10.

The data were derived from WoS. In WoS the phenomenon of linguistic bias is well-known (Frenken et al., 2009, Hoekman et al., 2010, Narin et al., 1991). The Anglo-Saxon countries are overrepresented in WoS. Hoekman, Frenken, and Tijssen (2010) analyzed the changing spatial patterns of scientific collaboration. Scientific collaboration is inhibited by linguistic differences among authors. This phenomenon creates a border among such co-authors. Hoekman, Frenken, and Tijssen (2010) found a strong effect of language border in research collaborations but in time it showed a decreasing tendency.

REFERENCES

Acosta, M., Coronado, D., Ferrándiz, E. & León, M.D. (2011). Factors affecting inter-regional academic scientific collaboration within Europe: the role of economic distance. *Scientometrics* 87, 63-74.

Barabási, A. L. (2014). Network science book. Network Science. pp. 17-19.

Bastian, M., Heymann, S., Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media.*

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies,* 39, 61-74.

Braun, T., Gómez, I., Méndez, A., & Schubert, A. (1992). International co-authorship patterns in physics and its subfields, 1981–1985. *Scientometrics*, 24(2), 181-200.

Csardi, G., Nepusz, T. (2006). The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. http://igraph.org

Csermely, P. (2005). *A rejtett hálózatok ereje*, Budapest, Vince kiadó: 376

Everett, M. G., & Borgatti, S. P. (2013). The dual-projection approach for two-mode networks. Social Networks, 35(2), 204-210.

Frenken, K., Hölzl, 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), 9-30.

Frenken, K., Hardeman, S., & Hoekman, J. (2009). Spatial scientometrics: Towards a cumulative research program. Journal of Informetrics, 3(3), 222-232.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, 178(4060), 471-479.

Goldfinch, S., Dale, T., & DeRouen Jr, K. (2003). Science from the periphery: Collaboration, networks and'Periphery Effects' in the citation of New Zealand Crown Research Institutes articles, 1995-2000. Scientometrics, 57(3), 321-337.

Hamers, L., Hemeryck, Y., Herweyers, G., Janssen, M., Keters, H., Rousseau, R., & Vanhoutte, A. (1989). Similarity measures in scientometric research: the Jaccard index versus Salton's cosine formula. Information Processing & Management, 25(3), 315-318.

Haustein, S. (2012). *Multidimensional journal evaluation: Analyzing scientific periodicals beyond the impact factor*. Walter de Gruyter. pp.223-252.

Hoekman, J., Frenken, K., & Tijssen, R. J. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy,* 39(5), 662-673.

Hou, H., Kretschmer, H., & Liu, Z. (2008). The structure of scientific collaboration networks in Scientometrics. *Scientometrics,* 75(2), 189-202.

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

Langfelder, P., Zhang, B., & Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, 24(5), 719-720.

Langfelder, P., Zhang, B. (2014). and with contributions from Steve Horvath (2014). dynamicTreeCut: Methods for detection of clusters in hierarchical clustering

dendrograms. R package version 1.62. https://CRAN.R-project.org/package=dynamicTreeCut

Leydesdorff, L., Carley, S., & Rafols, I. (2013). Global maps of science based on the new Web-of-Science categories. Scientometrics, 94(2), 589-593.

Luukkonen, T., Tijssen, R., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics*, 28(1), 15-36.

Meyer, D., Buchta, C. (2015). proxy: Distance and Similarity Measures. R package version 0.4-15. https://CRAN.R-project.org/package=proxy

Narin, F., Stevens, K., & Whitlow, E. (1991). Scientific co-operation in Europe and the citation of multinationally authored papers. *Scientometrics*, 21(3), 313-323

Nemes Nagy, J. (2009). *Terek, Helyek, Régiók. A regionális tudomány alapjai.* Budapest: Akadémiai Kiadó.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM review*, *45*(2), 167-256

Nguyen, H. V., & Bai, L. (2010). Cosine similarity metric learning for face verification. In *Computer Vision–ACCV 2010* (pp. 709-720). Springer Berlin Heidelberg.

Pudovkin, A. I., & Garfield, E. (2004). Rank-normalized impact factor: A way to compare journal performance across subject categories. *Proceedings of the American Society for Information Science and Technology*, *41*(1), 507-515.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Sen, B. K. (1992). Normalised impact factor. *Journal of Documentation*, *48*(3), 318-325.

Schubert, A., & Braun, T. (1990). International collaboration in the sciences 1981–1985. *Scientometrics*, 19(1-2), 3-10.

Wasserman, S., & Iacobucci, D. (1991). Statistical modelling of one-mode and two-mode networks: Simultaneous analysis of graphs and bipartite graphs. British Journal of Mathematical and Statistical Psychology, 44(1), 13-43.

http://www.businessdictionary.com/definition/collaboration.html

http://incites.isiknowledge.com/common/help/h_field_category_wos.html

http://wokinfo.com/products_tools/multidisciplinary/webofscience/?utm_source=false&utm_medium=false&utm_campaign=false

.