Research Article

Adelina-Alexandra Stoica*

Homophily in co-authorship networks

<https://doi.org/10.2478/irsr-2018-0014>

Received: July, 12, 2018; Accepted: July, 28, 2018

Abstract: The main purpose of this paper is to measure the impact that homophily, structural characteristics of the networks, number of citations of the alters and their Hirsch score have on the number of citations of an ego. I have chosen co-authorship networks as a subject of research because they have a great influence on knowledge and on the diffusion of ideas. The studied populations are represented by full-time academics affiliated to sociology departments in Romania, Poland and Slovenia. Ego-network analysis was used as research design. The data was analyzed using linear hierarchical regression. For all three populations the average number of citations of the alter has a considerable positive impact on the number of citations of the ego. Conversely, the Hirsch score of the alter has a negative impact on the number of citations of the ego. The data analyzed in this article claims that the assumptions about the positive impact of alter citations, network size and the betweenness score on the number of the authors citations are supported empirically.

Keywords: homophily, co-authorship networks, Hirsch score, structural features.

Introduction

Over the last decades, social network analysis has gained a great visibility, especially because this type of analysis has been applied interdisciplinary. The methodology associated with social network analysis was adopted in publications from various disciplines such as scientometry, bibliometry or biology (Rueda, Gerdri and Kocaoglu, 2007; Borgatti and Halgin, 2011; Matusiak and Morzy, 2012).

The use of social network analysis has had a substantial impact in studying the co-authoring process, the diffusion of ideas, understanding the formation of links between authors and understanding the formation of academic prestige (Barnett, Ault and Kaserman, 1988; Stokes and Hartley, 1989; Newman, 2004; Bozeman, Fay and Slade, 2013).

Several papers on co-authoring have been published in the literature. Following these researches, very important conclusions have been mentioned. For example, from Newman's research (2004), we find that the number of publications in the co-authoring varies depending on the field in which the researchers work. Another research (Hâncean and Perc, 2016) shows that the average of co-authors' citations has a positive impact on the number of citations of the ego.

An advantage for studying collaboration in science is to find out if and to what extent this phenomenon helps the increasing of the scientific productivity. Barnett, Ault and Kaserman (1988) used the co-authorship networks in articles as an effective proxy variable for measuring collaboration in science.

In this paper, the issue of academic impact measured by the h-index was also approached, since currently the h-index represents the most widely accepted way of measuring the academic impact of a publication, research group or journal (Engla, 2010).

The distance to which the ego is situated against the network was also an important aspect in the bibliometric studies (Hâncean and Perc, 2016) therefore I have chosen to include in my analysis the betweenness score seen as a factor that has a positive impact on the number of citations of the ego.

The main purpose of this paper is to explain the extent to which homophily and network features affect the number of individual citations of researchers that work full time in departments of sociology within universities in three different geographical areas: Slovenia, Poland and Romania.

*Corresponding author: Adelina-Alexandra Stoica,
stoicaalexandraadelina@gmail.com

Homophily

A social network is the representation of two or more nodes connected to each other, the connection criterion being the existence of a type of relationship between these nodes (Wasserman and Faust, 1994).

This paper focuses on ego-main networks. Ego-networks are composed of an ego, its alters, the relationship between the alters and the relationship between ego and the alters (Wasserman and Faust, 1994).

Homophily is the practice of interacting and associating with similar individuals (Fu et al., 2012). It is not only present in the social life of individuals. Its roots can be identified in the animal world by the tendency of animals to find partners as similar as possible (Sundaresan et al., 2007). On the opposite pole, heterophily, the tendency to interact with different individuals is also present in both the social and animal environment, this being less common than homophily (Fu et al., 2012). A context where heterophily is present and is of great benefit is the organizational environment where people tend to choose partners who have complementary skills (Fu et al., 2012). This behavior facilitates teamwork by combining strengths that positively lead to the completion of a task.

Fu et al., (2012) makes a distinction between homophily and heterophily based on the advantages and disadvantages of each concept. Among the benefits of homophily, it is important to note that it facilitates communication more effectively, since individuals are similar, creating synergy between them. When it comes to heterophily, an important advantage is the tendency and encouragement of individuals to specialize.

Another perspective from which we can look at homophily suggests that people tend to have relationships or connections with similar individuals and eventually this brings with it a lack of diversity in their view and the kind of information they have access to (McPherson, Smith-Lovin and Cook, 2001). Taking this argument into account, we can assume that homophily not only creates synergy but also social segregation.

Segregation refers to multiple aspects of the social environment. The important categories in which individuals are divided can also have negative consequences when they are present in an unfavorable context. For example, segregation in religion, race, gender or social class can turn into social problems. Its consequences are not always beneficial to both parties: see for instance, problems such as sexism, homophobia, religious, discrimination, and so on.

Lazarsfeld and Merton (1954) have detailed the concept of homophily using two categories based on

different aspects of this concept: value homophily and status homophily. In status homophily, it is explained that individuals tend to relate to individuals who are part of the same social class, have the same formal or informal social status, have the same sex, age, religion. Value homophily involves the relationship between individuals according to less visible features such as values, behaviors, beliefs and so on. Although this categorization has a different specificity, status homophily has a considerable weight in determining homophily based on values because the status has implications for values. For example, affiliation to a religion is closely linked to values through imposed rules. To be a good practitioner, you must have desirable behavior, behavior being a consequence of values (Dowling and Pfeffer, 1975).

Kandel (1978) shows that homophily is the process that mostly explains desirable and deviant behaviors. The results of the study (Kandel, 1978) show that for the explanation of behavior it is very important to look at how individuals choose their alters. Kandel (1978) mentions that individuals are choosing alters based on value-based homophily. This shows that the individual is prone to a certain behavior and they choose their group of friends according to this predisposition.

Homophily characterizes and determines to a considerable extent the shape of a network (McPherson, Smith-Lovin and Cook, 2001). Differences such as ethnicity, gender, age, social status give shape to the network we are part of through the personal choice of individuals regarding the introduction of alters into the personal network.

Mentioning the causes of homophily, McPherson, Smith-Lovin and Cook (2001) specify that along with the technology, the geographical distance is very important also. We tend to form relationships with people who are in the proximity of space rather than those to whom we are at a great distance. Although technology facilitates the formation of links with others by simplifying the process by which we connect with them, the spatial physical aspect remains relevant, being the first indicator to show how often we meet with our friends. In most cases, technology comes as support for relationships that have already been established in the past (McPherson, Smith-Lovin and Cook, 2001).

Urban areas have a peculiarity in determining relationships. By their characteristic of not having a very large geographic space but hosting many residents, this context creates relationships based on this type of segregation regarding race and ethnicity (Marsden, 1987).

Another context that facilitates the creation of links is the formal education organization. Schools and

universities not only bring individuals together and give them the opportunity to create such links, but also helps in maintaining the links by requiring individuals to physically partake in mutual activities such as classes over long periods of time (McPherson, Smith-Lovin and Cook, 2001).

Using data from three databases on three disciplines: biology, physics and mathematics, Newman (2004) builds networks where the nodes are researchers and the connecting criterion is the presence of an article written together. The author looks into detail at three collaborative networks. The results show that in all three of these areas the number of co-authors varies a great deal from several authors to “hundreds or even thousands in some cases”. Researchers working in the field of biology have a significantly higher number of co-authors than in the other two areas. An explanation for this is the direction towards the experiment method, a context that creates co-authors because of the great workload that has been made in carrying out experiments (Newman, 2004). Other discoveries are harder to explain. For example, in biology, it is less likely than in mathematics that two co-authors will co-author in other future articles (Newman, 2004).

The article on which this papers approach and data are based is a comparative research carried out by Hâncianu and Perc (2016) that explains to what extent and how homophily and network properties affect the number of citations in the sociology departments in three countries: Poland, Romania and Slovenia.

Empirical data on three populations of researchers was used. The principle of choosing nodes is the principle of structural equivalence. The researchers that were analyzed working full time in the sociology departments of universities. The dependent variable is the number of citations of the ego, and the independent variables are: the number of publications, the average score of the citations of the other publications, the betweenness score, the density and the size of the network.

The results show that authors in Slovenia are part of wider networks of co-authoring than in Romania and Poland, which facilitates the transmission of information within the network. In the case of Romania, the authors are part of smaller but more dense networks. The hypothesis stating that the average of the citations of a researcher's co-authors has a significant impact on the researcher's citations is supported empirically. Another hypothesis suggesting that the size of the network has a positive impact on authors' citations has been supported empirically, but only for Slovenia and Poland.

Bibliometric analyzes can be very useful because they show how fragmented a network is or what is the degree

of cohesion in a scientific community. Another important idea that can be highlighted by this kind of analysis is the observation of the best connected actors, which play an important role in the co-authorship networks.

Structural characteristics of networks

The structural features of a network refer to the type of links present between pairs of actors (Wasserman and Faust, 1994). Structural variables are variables that refer to the results derived from the links between two or more actors (Wasserman and Faust, 1994).

The study of co-authors from a network-based perspective can show that if researchers change ideas, goals, research questions, validation methods, similar methods of data analysis, then these cohesive networks can generate consensus on issues, research methods and results (Friedkin, 1998).

Moody (2004) promotes the idea that collaboration is unequally divided. Because a small number of researchers are the key actors of a network, other researchers have access to different science specialized clusters through these very active individuals within the network (Moody, 2004). This advantageous position shows why key actors have the ability to be more influential and how these researchers have the opportunity to diffuse their ideas in the scientific community (Moody, 2004). Due to homophily, key researchers are inclined to collaborate with other researchers with the same status feature (Moody, 2004). The integration of new ones into such networks depends heavily on similarity to the ideas of the key actors and their collaborators as they follow the theoretical trajectory promoted by the key actors (Barabási and Albert, 1999).

A type of structure that more efficiently supports collaboration in science is the cohesive network from a structural point of view. In contrast to the preferential attachment structure, the cohesive network links are distributed in such a way that key actors no longer have a very important role in the network (Moody, 2004).

Moody and White (2003) define the cohesive structure network as a community where relations between these members maintain the network. Unlike the preferential attachment network where if the key actors disappear, the structure of the network also disappears, within the cohesive structure links are evenly distributed, others having access to each other, not depending on the key actors.

By analyzing co-authoring networks through the collaboration structure, Newman (2001) shows that these

networks are very similar to small-world networks where researchers are separated by a very low geodesic distance from intermediate researchers. The author shows that these networks follow the power law (e.g. preferential attachment).

Co-authoring and citations networks

An important feature of papers with original content, especially the ones that are creating data through the experimental method, is the presence of co-authors (Stokes and Hartley, 1989). Through the analysis of the co-authors, one can observe the main papers in the field, but not the central authors. This is due to the fact that the order of authors is not an indicator of confidence in determining the contribution of each author.

In order to better exemplify the co-authorship ego-network, I propose Figure 1. The network was built in UCINET 6 software (version 6.616) (Borgatti, Everett & Freeman, 2002)).

The network of Figure 1 is a random ego-network, with the sole purpose of using it as an example. Nodes are regarded as authors and the links between them show that the nodes participated in writing an article together. The size of the nodes shows the hierarchy according to the

number of co-authors. We notice that nodes 14, 11, 7 and 5 have the most co-authors in the network. Nodes 4, 6, 8 and 12 are isolated nodes that did not co-authored.

Citation score is one of the most commonly used metric in bibliometry (Stokes and Hartley, 1989). Citation analysis has served as a tool to determine the scientific impact. Citations are also used as an indicator for awarding research funds because they are designed to assess how performant a researcher is (Bauer and Bakkalbasi, 2005).

It is already accepted that papers by several authors are not the only channel of communication between co-authors. Formal communication through articles is rather secondary, the main mean of communication being of an informal nature, where researchers in the same field change ideas (Price, 1963).

Citation networks are considered a type of social network themselves. Because studies that have as a research object the citations, have used statistical-based methods, they have been studied rather in a quantitative way, helping to develop databases that capture information about them. The most representative indexes on this subject are used in scientometry, being defined as a set of techniques and tools used in quantitative analysis and in the measurement of the scientific impact of publications (Price, 1963). At the same time citations are used to assess the quality of scientific publications and analyze the development of various areas of science.

Figure 1. Co-authorship ego-network

Hirsch score

The evolution of scientometry fundamentally depends not only on the degree of specialization of the researchers dealing with this subject, but also on the methods and tools used to research this subject (Garfield, 1970).

The co-authoring network, in addition to the transmission of information and the creation of international research communities also has associated disadvantages. Since writing articles is a prerequisite for belonging to an academic community, it raises a major problem regarding the quality of the works and the ways in which quality can be measured. To address this challenge, scientometry, the science of measuring and analyzing science, provides an answer. Scientometry includes different types of measurements, many of which are constructed on the number of publications and citations (Garfield, 1970).

The name of the score comes from Argentinean physicist Jorge Hirsch, physics professor at the University of San Diego, USA.

Jorge Hirsch is a character of the academic scene that sparked great reactions in the academic world. His purpose was to respond to the need to find a suitable measure to accurately assess the academic achievements of scholars. To this challenge, Hirsch responds by creating a measurement system that measures a researcher's performance through an evaluation criterion based on the number of citations. A researcher has the h-index only if a number of "h" scientific articles of the total articles published by him have at least a "h" number of citations (Hirsch, 2005). The higher the score associated with a researcher, the more a researcher's performance is valued in the academic world.

Since the number of papers does not reveal in any way the quality of the works, assigning a value has become an important measure in measuring quality (Garfield, 1970). Various quality measurement indices have been introduced, such as score g or Hirsch score. Although there is controversy in the academic world about the efficiency with which the h-index measures the quality of a work, it is still used to measure academic performance.

For a set of works hierarchized within the number of citations the author has received, the h index is the highest number that received *h* or more quotes (Engla, 2010). The h-index has the advantage that it is a number that manages to incorporate both the quantitative part (the number of publications) and the qualitative part referring to the visibility (the number of citations) (Engla, 2010).

Another advantage is that the Hirsch score is not sensitive to sets of not cited or slightly cited publications

or to papers that have a much higher number of citations compared to the other publications (Engla, 2010). While the author agrees with the advantage of the insensitiveness about non cited or slightly cited publications, he considers that insensitivity to the large number of citations can be repaired because the Hirsch score does not take into account the continuity of citations over time.

Application of the Hirsch index does not only focus on the achievements of a single individual but can also be applied to groups of researchers, universities or journals. From the micro level it goes to the macro level. The Hirsch score also helps with its ability to be compared with another Hirsch index in the same field (Bornmann and Daniel, 2007).

Bornmann and Daniel (2007) brought into attention several shortcomings of the Hirsch score. The first one refers to the fact that using h-index is not beneficial for a researcher's academic work because it is limited to a single descriptive figure and it forces the multidimensionality of bibliometry to be reduced to a single equivalent dimension of a figure.

Another shortcoming is the incapacity of the h-index to distinguish between active and inactive or outgoing researchers, but also the inability of the score to distinguish between significant past work and papers focused on highly popular subjects or papers with an influence on topics addressed by the community (Bornmann and Daniel, 2007).

Because the Hirsch score also depend on the cumulative time a researcher has worked in the field through publications, this puts an alarming signal on young researchers who start with a disadvantage compared to experienced researchers who are already a part of the scientific community (Bornmann and Daniel, 2007).

After studying the literature on the Hirsch score, Bornmann and Daniel (2007) conclude by mentioning that the h-index is a good tool for validating the researchers' performance at micro and meso level.

Hirsch recognizes the limits of his own score, thus ruling that the score can not measure all aspects of the scientific impact alone. The researcher recommends the use of the h index with other forms of scientific impact measurement when quantifying academic results (Hirsch, 2005).

In order to minimize the disadvantageous effects on young researchers, Hirsch (2005) introduces the *m* parameter, a measure dividing the Hirsch score taking into account the number of years passed from the first published work of an author.

The use of the index h leads to the rewarding of the quantity in exchange for the impact of the publications (Costas and Bordons, 2007). As the authors illustrate, when a researcher who has ten publications is cited ten times and a researcher B has five publications but is cited 200 times, the Hirsch score of researcher A is ten and Hirsch score of researcher B is five, the Hirsch score no longer responds to the need to measure the impact of the publication. The resulting question in this example is: Will researcher A be regarded as having greater success than researcher B? Unfortunately, if we only consider the value of the index h , the answer is yes.

Another situation where the use of the h -index can raise concerns in academia is where researchers have the same Hirsch score but a very different number of citations (Costas and Bordons, 2007). An edifying example demonstrating the lack of infallibility of this score is when two authors publishing five documents have the same score, although one has ten citations per document, while the other has twenty citations per document (Costas and Bordons, 2007). This raises doubts as to the comparability of these two researchers in the context of scientific impact because the h -index tends to underestimate authors with a small number of publications but of great international impact.

Method and data

Predictors used to explain the dependent variable, the number of citations of the ego are: the average of the number of citations of the alter, the average of the Hirsch score of the co-authors, the score betweenness and the size of the network.

Predictors used were chosen from similar works. These predictors have been included in a variety of papers addressing the study of co-author writing (Stokes and Hartley, 1989; Moody, 2004; Newman, 2004; Hâncean and Perc, 2016).

The first hypothesis refers to the average number of citations of the alters that will have a positive impact on the ego citations. The second hypothesis is about the size of an egos' network that has a positive impact on the number of citations. The third hypothesis refers to Hirsch's score of co-authors that will have a positive impact on the number of citations of the author. The last hypothesis states that the higher the betweenness score, the more citations an ego has.

A very important aspect to be mentioned is that in the statistical analysis every variable used was logarithmized because the variables contained extreme cases observed

by histograms and scatterplots, the logarithm was necessary to uniform their distribution. I did not exclude the extreme cases from the analysis because, although it increases the distribution that would otherwise have been normal, these cases reflect a part of the social reality captured by the data.

Data collection

The data included three countries: Slovenia, Poland and Romania. The motivation for this choice is the curiosity about how knowledge is being spread and the impact of homophily within the co-authorship in different geographical areas. For the countries chosen to be comparable, the choice of countries was made based on the historical context. Slovenia left the communist political infrastructure in 1990, Poland in 1989 and Romania also in 1989.

The database used to analyze the impact of homophily and network characteristics on the number of individual citations is the database used by Hâncean and Perc in the article *Homophily in co-authorship networks of East European sociologists* (2016) (Hâncean, 2016).

In January 2016, the population in the two sociology departments of the Universities from Slovenia (Ljubljana and Maribor) numbered 58 full-time academic sociologists, the population in the 10 sociology departments of the Polish Universities was represented by 55 full-time academic sociologists and in Romania, the population was 294 belonging to the 16 sociology departments.

The database was built using the Web of Science platform, a bibliographic and bibliometric database available from Tomson Reuters, which indexes approximately 9,000 journals.

The database also includes attributes such as: departmental affiliation, number of publications, number of co-authors, number of citations, h -index of each researcher.

Data on departments and departmental researchers were taken from official university sites.

Data analysis

To test the variation in the number of ego publications I have used linear hierarchical regressions. I then built five regression models by gradually adding independent variables to the model to see how much of the dependent variable is explained by predictors for each model.

To observe whether the hypothesis are supported empirically, I built linear hierarchical regression models. An advantage that the linear hierarchical regression

offers is also the specification of the explanatory capacity difference of the theoretical model with the addition of the independent variables gradually. Through this statistical operation we can see which predictors hold the greatest weight in explaining the variation in the number of quotes of the ego.

For population descriptions I chose to include statistical operations such as: average, minimum value, maximum value, and standard deviation.

Results

Romania

In order to form a general idea of co-authorship, I described the general situation regarding the data-derived co-authorship with descriptive statistics.

By comparing the average scores and the standard deviation in Table one, we can see that there are large differences between the authors regarding the average number of citations of the alter. If we look at the average size of the network and the standard deviation, we find that there are only three standard deviations above the average, which shows that the authors are quite similar in this area and it indicates the presence of homophily.

From table number two, we can notice that step 3 explains the most effective variation in the number of citations of the ego. Predictors are the average of the citations, the Hirsch score of the alter and the normalized betweenness score. They manage to explain 61% of the ego citations.

In Table number two we can see the power with which theoretical models explain the ego's citations number. Although a predictor is added in step four, its impact decreases the ability to explain by 0.4%. We note that by introducing the mean Hirsch score of co-authors and the betweenness score, the model manages to explain 3.4% more than step 2.

Poland

From table number three, we can deduce that there is homophily at the Hirsch score level of the alters because the standard deviation is not very distant from the mean value. Another variable that seems to support the idea of homophily among individuals is the normalized betweenness score because there are only 0.2 standard deviations above the average.

For both Romania and Poland, the first hypothesis is supported by data at least for Romania where the

standardized Beta coefficient is substantially higher. Although in the case of Poland the Beta coefficient does not have the same magnitude, it should not be overlooked.

As with Romania, model 3 best explains the variation of the dependent variable. Unlike the case of Romania, where the model explains 61% of the dependent variable, in the case of Poland, the model manages to explain only 13% of the ego citations.

In Table 4 we can observe the ability of models to explain the variation of the dependent variable. Although a predictor is added in model four, its impact decreases the ability to explain by 3%. Step 3 manages to explain most of the variation of the ego's citations number.

Table 1. Descriptive statistics (Romania)

Means, standard deviation, minimal and maximal values of the variables					
Variable	N	Mean	Min	Max	S.D.
Ego's citations number	120	7	0	185	25.5
Mean of co-authors' h index	93	1.6	0	26	3.2
Mean of alters' citations	93	82.5	0	4079.5	433.9
Normalized betweenness	93	.1	0	.9	.2
Size of the network	120	3	0	20	3.5

Table 2. Hierarchical regression models (Romania)

Variable	Step 1	Step 2	Step 3	Step 4
1 Mean of alters' citations	.764	1.371	1.176	1.169
2 Mean of co-authors' h-index		-.611	-.530	-.524
3 Normalized betweenness score			.224	.220
4 Size of the network				.006
R ²	.580	.581	.615	.611
ΔR ²		.001	.034	-.004

Table 3. Descriptive statistics (Poland)

Means, standard deviation, minimal and maximal values of the variables					
Variable	N	Mean	Min	Max	S.D.
Ego's citations number	55	11.1	0	474	64
Mean of co-authors' h-index	31	1.3	0	5.5	1.4
Mean of alters' citations	31	55.9	0	472	102.7
Normalized betweenness	31	.1	0	.9	.2
Size of the network	55	2.4	0	28	5

Slovenia

From Table 5 we note that there is not a large variation between the mean of co-authors' h-index in the Slovenian population studied, result that indicates the presence of homophily. The table also shows significant differences between the ego's citations number. In Slovenia, the standard deviation of the mean of alters' citations is 55 above the average. This shows the lack of homophily in this specific context.

The variation in the number of citations of the ego is best explained by the theoretical model number four (see table number six). We note that the average number of citations of the alter, the betweenness score and the size of the network have a positive impact on the dependent variable. Predictors account for 46% of the ego citations.

In Table 6 we can also see the power with which theoretical models explain the dependent variable. Although a predictor is added to model four, its impact improves the ability to explain only by 1.2%. We note that by introducing the average h scoring coefficients of the co-authors and the betweenness score, the model manages to explain 9.3% more than model 2.

Discussion and conclusions

This paper started from the assumption that homophily is part of the formation, structure and composition of social networks (Kandel, 1978; McPherson, Smith-Lovin and Cook, 2001; Fu et al., 2012; Boucher, 2015). This principle also applies to the networks of co-authorship (Hâncean and Perc, 2016). In the continuation of these researches, the present paper also tests the homophily effect. The hypothesis that the average number of citations of the alter has a positive impact on the number of quotes of the ego is supported empirically. This indicates the presence of homophily in the networks.

The results of this paper show that alters' citations has a positive impact upon the number of ego's citations. Although some researchers have noted that the co-authors' h-index has a positive impact on the variation of the ego citations (Stokes and Hartley, 1989; Newman, 2004), the results of this paper show that the co-authors' h-index has a rather negative impact on the number of citations of the ego in all three countries when the mean of alters' citations is also present in the models.

In Romania, when the co-authors' h-index grows with a standard deviation, the number of citations of the ego is decreases by .53 standard deviations. In Poland by .64 standard deviations and in Slovenia when the co-authors'

Table 4. Hierarchical regression models (Poland)

Variable	Step 1	Step 2	Step 3	Step 4
1 Mean of alters' citations	.372	.934	.912	.922
2 Mean of co-authors' h index		-.572	-.648	-.681
3 Normalized betweenness score			.280	.213
4 Size of the network				.094
R ²	.109	.090	.133	.103
ΔR^2		-.019	.043	-.03

Table 5. Descriptive statistics (Slovenia)

Means, standard deviation, minimal and maximal values of the variables					
Variable	N	Mean	Min	Max	S.D.
Ego's citations number	58	31.7	0	284	50.2
Mean of co-authors' h-index	50	2.4	0	6	1.2
Mean of alters' citations	50	47.8	0	321.5	55.4
Normalized betweenness	50	4.5	0	25.8	6.4
Size of the network	58	4.7	0	29	4.7

Table 6. Hierarchical regression models (Slovenia)

Variable	Step 1	Step 2	Step 3	Step 4
1 Mean of alters' citations	.542	1.654	1.432	1.199
2 Mean of co-authors' h-index		-1.152	-1.047	-.843
3 Normalized betweenness score			.340	.214
4 Size of the network				.216
R ²	.279	.356	.449	.461
ΔR^2		.077	.093	.012

h-index increases with a standard deviation, the number of ego quotes decreases by .84 standard deviations.

The structural characteristics like betweenness score and the size of the network have a positive impact upon the citations number of an ego in Romania, Poland and Slovenia. In Romania, when the betweenness score and the size of the network increases with a standard deviation, the citations number of the ego increases with .224 and .006 standard deviations. In Poland when these structural characteristics increase with one standard deviation, the ego's citations number increases by .280 and .094 standard deviations, in Slovenia with .214 and .216. The results show that in Romania and Poland, from the category of structural characteristics, the size of the network has the lowest impact. In Slovenia, the standardized Beta coefficients

are similar for betweenness score and for the size of the network.

Following the observation of the coefficient that explains how much percent of the dependent variable is explained by the independent variables we can conclude that the co-authoring process is different because in the case of Romania and Poland, using the same theoretical model, it explains in different proportions the variation in the number of citations of the ego (Romania = 61%, Poland = 13%). In Slovenia, the variance of the dependent variable is explained by 46% of all the independent variables included in the research.

One limit of research is that this paper only approached quantitative methods. I believe that a mixed approach involving both quantitative methods and qualitative methods would have had a positive impact on explaining more of the variation of the ego's citations.

A shortcoming of this research is that the database did not include the gender of the co-authors. An analysis that also takes into account the gender of the alter brings additional knowledge because in this way one can resort to gender theories, and thus more complex explanatory patterns can be noticed which can explain the variation in the number of citations of an author. Another factor of the research that should be considered is the academic rank of the studied population. This addition would help explain the overall image involving homophily because that would show if the academic rank is a variable that counts in the decision of choosing a coauthor.

In the future, I would like my research to curb all of the above-mentioned limits, and I also propose to address this topic from a dynamic and longitudinal perspective. I believe that a longitudinal approach brings a great deal of knowledge because we can observe the evolution of these populations over time. In this paper there were not presented issues related to the visualization of co-authoring networks in the paper. This aspect could give us a better image of the homophily in the co-authoring networks.

Acknowledgements: This research was funded through the project “*Longitudinal analysis of co-authorship networks and citations in academia*” (iCoNiC), funded by the Executive Unit for Financing Higher Education, Research, Development and Innovation (UEFISCDI) - code: PN-III-P1-1.1-TE-2016-0362.

References

Barabási, A. L. and Albert, R. (1999) ‘Emergence of scaling in random networks’, *Science*, 286(5439), pp. 509–512. doi: 10.1126/science.286.5439.509.

- Barnett, A., W Ault, R. and L Kaserman, D. (1988) ‘The Rising Incidence of Co-Authorship in Economics: Further Evidence’, *The Review of Economics and Statistics*, 70, pp. 539–543.
- Bauer, K. and Bakkalbasi, N. (2005) ‘An examination of citation counts in a new scholarly communication environment’, *D-Lib Magazine*. doi: 10.1045/september2005-bauer.
- Borgatti, S. P. and Halgin, D. S. (2011) ‘Organization Science’, *Organization Science*, 22(5), pp. 1168–1181.
- Bornmann, L. and Daniel, H. D. (2007) ‘What do we know about the h index?’, *Journal of the American Society for Information Science and Technology*, 58(9), pp. 1381–1385. doi: 10.1002/asi.20609.
- Boucher, V. (2015) ‘Structural homophily’, *International Economic Review*, 56(1), pp. 235–264. doi: 10.1111/iere.12101.
- Bozeman, B., Fay, D. and Slade, C. P. (2013) ‘Research collaboration in universities and academic entrepreneurship: The-state-of-the-art’, *Journal of Technology Transfer*, pp. 1–67. doi: 10.1007/s10961-012-9281-8.
- Costas, R. and Bordons, M. (2007) ‘The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level’, *Journal of Informetrics*, 1(3), pp. 193–203. doi: 10.1016/j.joi.2007.02.001.
- Dowling, J. and Pfeffer, J. (1975) ‘Organizational Legitimacy: Social Values and Organizational Behavior’, *The Pacific Sociological Review*, 18(1), pp. 122–136. doi: 10.2307/1388226.
- Egghe, L. (2006) ‘An improvement of the h-index: The g-index’, *ISSI Newsletter*, pp. 1–4.
- Engla, N. E. W. (2010) ‘New england journal’, *Perspective*, 363(1), pp. 1–3. doi: 10.1056/NEJMp1002530.
- Friedkin, N. E. (1998) ‘Toward a Structural Social Psychology’, in A *Structural Theory of Social Influence*, pp. 23–34.
- Fu, F., Nowak, M. A., Christakis, N. A. and Fowler, J. H. (2012) ‘The evolution of homophily’, *Scientific Reports*, 2. doi: 10.1038/srep00845.
- Garfield, E. (1970) ‘Citation indexing for studying science’, *Nature*, 227(5259), pp. 669–671. doi: 10.1038/227669a0.
- Hâncean, M. (2016) ‘AQuaRel’. România.
- Hâncean, M. G. and Perc, M. (2016) ‘Homophily in co-authorship networks of East European sociologists’, *Scientific Reports*, 6. doi: 10.1038/srep36152.
- Hirsch, J. E. (2005) ‘An index to quantify an individual’s scientific research output’, *Proceedings of the National Academy of Sciences, U.S.A.*, 102(46), pp. 16569–16572. doi: 10.1073/pnas.0507655102.
- Kandel, D. B. (1978) ‘Homophily, Selection, and Socialization in Adolescent Friendships’, *American Journal of Sociology*, 84(2), pp. 427–436. doi: 10.1086/226792.
- Lazarsfeld, P. F. and Merton, R. K. (1954) ‘Friendship as a Social Process: A Substantive and Methodological Analysis’, in *Freedom and Control in Modern Society*.
- Marsden, P. V. (1987) ‘Core Discussion Networks of Americans’, *American Sociological Review*, 52(1), p. 122. doi: 10.2307/2095397.
- Matusiak, A. and Morzy, M. (2012) ‘Social Network Analysis in Scientometrics’, *2012 Eighth International Conference on Signal Image Technology and Internet Based Systems*, pp. 692–699. doi: 10.1109/SITIS.2012.105.
- McPherson, M., Smith-Lovin, L. and Cook, J. M. (2001) ‘Birds of a Feather: Homophily in Social Networks’, *Annual Review of Sociology*, 27(1), pp. 415–444. doi: 10.1146/annurev.soc.27.1.415.

- Moody, J. (2004) 'The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999', *American Sociological Review*, 69(2), pp. 213–238. doi: 10.1177/000312240406900204.
- Moody, J. and White, D. R. (2003) 'Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups', *American Sociological Review*, 68(1), p. 103. doi: 10.2307/3088904.
- Newman, M. E. J. (2001) 'Scientific collaboration networks. I. Network construction and fundamental results', *Physical Review E*, 64(1), p. 016131. doi: 10.1103/PhysRevE.64.016131.
- Newman, M. E. J. (2004) 'Co-authorship networks and patterns of scientific collaboration', *Proceedings of the National Academy of Sciences*, 101(Supplement 1), pp. 5200–5205. doi: 10.1073/pnas.0307545100.
- Price, D. J. D. S. (1963) *Little Science, Big Science, Little Science, Big Science*.
- Rueda, G., Gerdtsri, P. and Kocaoglu, D. F. (2007) 'Bibliometrics and social network analysis of the nanotechnology field', *Portland International Conference on Management of Engineering and Technology*, pp. 2905–2911. doi: 10.1109/PICMET.2007.4349633.
- Stokes, T. D. and Hartley, J. A. (1989) 'Co-authorship, Social Structure and Influence Within Specialties', *Social Studies of Science*, 19(1), pp. 101–125. doi: 10.1177/030631289019001003.
- Sundaresan, S. R., Fischhoff, I. R., Dushoff, J. and Rubenstein, D. I. (2007) 'Network metrics reveal differences in social organization between two fission-fusion species, Grevy's zebra and onager', *Oecologia*, 151(1), pp. 140–149. doi: 10.1007/s00442-006-0553-6.
- Wasserman, S. and Faust, K. (1994) *Social network analysis : methods and applications*, *American Ethnologist*. doi: 10.1525/ae.1997.24.1.219.