Social Network Analysis:
Bibliographic Network Analysis of the Field and its Evolution
Part 2. Analysis of Co-occurence Networks

Daria Maltseva[1], Vladimir Batagelj[1,2,3]

[1]NRU HSE Moscow
[2]IMFM Ljubljana
[3]IAM UP Koper
d_malceva@mail.ru
vladimir.batagelj@fmf.uni-lj.si

December 22, 2018

Abstract

Abstract ...

Keywords: social network analysis, bibliographic networks, main path, island,

1 Introduction

This paper presents the second part of the results of the study on the development of social network analysis (SNA) discipline and its evolution over time. We aim to implement comprehensive approach for the identification of the current trends of the Social network analysis field development – in terms of the representation of various disciplinary areas, groups of scientists, and thematic agenda in the field. The provided methodology of **bibliometric networks analysis** already proved to be productive in a set of studies of different scientific fields and topics (Kejžar et al., 2010; Batagelj et al., 2014, 2017).It allows allocating key publications and actors (authors, research groups, institutions, journals) in the SNA field, main topics and scientific ideas, connections between them and their evolution through time, as well as analyzing networks of co-authorship, co-occurence, citation and co-citation between different bibliographic entities.

The development of the SNA field was reflected in a set of studies focused both on its historiographical description (Freeman, 2004) and bibliometric analysis of publications and journals involved to the field (including *Social Networks*). Different authors studied citation structures of works and journals (Hummon and Carley, 1993; Leydesdorff et al., 2008; Batagelj et al., 2014), collaboration structures in sense of co-authorship (Otte and Rousseau, 2002; Leydesdorff et al., 2008; Batagelj et al., 2014), structures of co-citations between works, authors, and journals (Brandes and Pich, 2011), topical structures and keyword co-occurence networks (Leydesdorff et al., 2008; Groenewegen et al., 2015). Attention was also given to different subfields (subtopics) (Hummon et al., 1990; Kejžar et al., 2010; Batagelj et al., 2014, 2019) and subdisciplines within the field (Otte and Rousseau, 2002; Borgatti, Foster, 2003; Lazer et al., 2009; Varga, Nemeslaki, 2012). We presented the comrehensive review of these studies in the first paper [first paper citation].

Our dataset consists of articles from the *Web of Science* Clarivate Analytics database (Core Collection) and those published in the main journals in the field, created by searching for the key word "social network*." Using **WoS2Pajek 1.5** (Batagelj, 2017), we transformed our data into a collection of networks: one-mode citation network **Cite** on works (from the field CR of WoS file description) and two-mode networks – the authorship network **WA** on works $\times$ authors (from the field AU), the journalship network **WJ** on works $\times$ journals (from the field CR or J9), and the keywordship network **WK** on works $\times$ keywords (from the fields ID, DE or TI). An important property of all these networks is that they share the same first node set – i.e. the set of works (papers, reports, books, etc.) – wich means that they are *linked* and can be easily combined using the network multiplication into new *derived* networks (Batagelj et al., 2014). Works that appear in descriptions can be of two types: those which have full descriptions (*hits*), and those which were only cited (listed in the CR fields, but not contained in the hits). the cited only works ($DC = 0$) only partial descriptions are provided: we have information only about the *first* author, the journal and the publication year, and we have no information on the keywords (as there are no titles in ISI names and cited works). That is why for further analysis we constructed networks, which contain only works with complete description ($DC > 0$). We labeled these *reduced networks* **CiteR**, **WAr**, **WJr**, and **WKr**. In obtained networks, the sizes of sets are as follows: works $|W| = 70,792$, authors $|A| = 93,011$, journals $|J| = 8,943$, key words $|K| = 32,409$. The procedure of data collection, cleaning and networks construction is presented in detail in the first paper [citation].

In the first paper, we presented the analysis of basic networks **Cite**, **WA**, **WK**, and **WJ** (and their reduced versions), and thus extracted the most cited works, authors and journals with the largest amount of works, and most often used keywords in the SNA field. Using the Search path count approach, we extracted the main path, key-route paths and link islands in the citation network **Cite**. Based on the probabilistic flow node values, we identified the most important articles. The results show that starting from its institutionalization in the 1980-1990's, SNA field has grown significantly in terms of the number of publications and the amount of disciplines involved into the research using SNA approach. The number of publications shows the constant growth, and on average it doubles every 3 years. The analysis confirmed the previous studies on the SNA field development using citation network analysis. Up to the middle of 1990's the most "important" works belong to the authors from the *social sciences*, and starting from 2000's the field experience the "invasion of physicists". To our surprise, from 2010's both groups experience the "invasion" of scientists from a completely another field – *animal SNA*. According to the analysis of journals, another active field of SNA research goes from the *Computer science* field.

The obtained results put new question to the research. From one side, the lists of most cited works, most used journals and, especially, keywords (with top words *social, network* and *analysis*) do not contradict our basic knowledge of the SNA field, and thats why a conclusion on the relevance of the obtained data to the research objects can be done. From another side, the citation network analysis was not able to detect some well known topic groups from the field of social science, who has been very active recently, such as *probabilistic approach* with Snijders and Robins as representatives, *signed networks* developed by Doreian, etc. The explanation of this situation may lie in the nature of the **Cite** citation network and algorithms of **Main path** and **Islands** used for identification of main subgroups from it.

That's why in the current paper we answer to the posed questions using more complex analysis of other available networks, derieved from the basic networks with the procedures of **networks multiplication** and using **fractional approach** for normalization. In the following parts, we provide the description of the *topic structure of the field* based on the keywords co-occurence netwok analysis, the structures of *authors collaboration* and keywords associated with coathorship islands, patterns of *citation and co-citation* among authors and journals.

2 Data

2.1 Derieved networks

Using basic networks **CiteN**, **WAn**, **WJn**, **WKn** and their reduced versions **CiteR**, **WAr**, **WJr**, **WKr** we constructed other networks for the further analysis using the procedure of **networks multiplication**. Two-mode networks are composed of descrete two-mode arrays, which can be combined using matrix multiplication. The product of two compatible networks is the network corresponding to the product of matrices corresponding to the given networks. Two two-mode networks are *compatible* for multiplication, if the second set of vertices in the first network is equal to the first set of vertices in the second network. If all weights in two two-mode networks are equal to 1, then the product of the weights multiplication will also be equal to 1.(Batagelj and Cerinšek, 2013; Batagelj et al., 2014).

In our case, this shared set is a set of works (papers, reports, books, etc.), which *links* different basic and reduced networks to each other. Using multiplication, we constructed *derieved networks* of two types. First type are one-mode networks made by the multiplication of two two-mode networks. Multiplying *same* two-mode networks, we got the network of keywords co-occurence **KK** (**WKT** * **WK**) and collaboration network **AA** (**WAT** * **WA**). The weight of link between two nodes *w(k1,k2)* in keywords co-occurence **KK** network shows how many times the keywords *k1* and *k2* were used together in the same work. The wight of edge between two authors *w(u,v)* in collaboration network **AA** shows the number of works to which *u* and *v* both contributed. Multiplying *different* two-mode networks (which still share the same set of nodes), we constructed the network of authors and keywods **AK** (**WAT** and **WK**) counting in how many works the author *u* used the keyword *k*.

Another type of networks are those which are produced by three steps of multiplication, when we use some network, such as citation network, for a "projection" of relations between other bibliographic entities. Multiplying **WA** networks through the **Cite** network, we got the network of citations among authors **CiteA**, where the weight of edge between two authors *w(u,v)* shows the number of times when author *u* cited author *v*. Multiplying **WJ** networks through the **Cite** network we constructed the network of citations among journals **CiteJ**, where the weight of edge between two journals *w(i,j)* shows the number of times when journal *i* cited journal *j*. Similarly, out of **WA**, **WJ** and **Cite** networks we constructed the networks of co-citations among authors **ACoj** and and journals **JCoj**, where the weight of edge between two nodes (authors or journals) shows the similarity of their citation patterns.

The detailed description on each derieved network construction is presented in the corresponding sections.

2.2 Normalized networks

It was shown that the multiplication has some limitations, such as the overrating of the contribution of bibliographic entities with many ties (works having a lot of authors or keywords, journals having a lot of works). That's why the **fractional approach** (Gauffriau et al. 2007, Batagelj and Cerinšek (2013)) was proposed, which takes into account the contribution of bibliographic entities (works, authors, or journals), normalizing the weights between them in such a way their input is equal to 1.

Let us provide the example of authorship two-mode network **WA**. In regular network, the outdegree is equal to the number of authors of this work, and the indegree is equal to the number of works to which author contributed. The normalization create network $n(\mathbf{WA})$ where the weight of each arc is divided by the sum of weights of all arcs having the same initial node as this arc (outdegree or outsum of a node). A contribution of each paper p was equal to 1, and we assume that each author contributed to the work equally.

$$n(\mathbf{WA})[w,a] = \frac{\mathbf{WA}[w,a]}{\max(1, \text{outdeg}(w))}$$

A similar normalization of collaboration links, but with *outdegree(w) – 1* instead of *outdegree(w)* was proposed by Newman (2001), who interpreted the weight of link between two authors as a proportion of time spent for the collaboration with each co-author (that's why "collaboration" with herself or himself is not taken into account).

The proposed ways of normalization can be used in other two-mode networks. The detailed description of their usage for networks creation is presented in the corresponding sections.

3 Keywords co-occurence: Topic structure of the field

3.1 Network KKn production

To construct the one-mode network **KKn**, we normalized the reduced **WKr** network in such a way that all works have equal contribution by keywords they provide: the fractional contribution of each complete subgruph in KKn is equal to 1. With this normalization, the works with large number of keywords are not overrepresented – the inputs of works having 1 and 10 keywords are the same. Normalization creates network $n(WKr)$ where the weight of each arc [w, k] is divided by the sum of weights of all arcs having the same initial node (works) as this arc (outdegree of a node). The normalized network WKn was transposed and multiplied with itself. In the obtained network, the loops were deleted and bidirected arcts were transformed to edges (with summation of the line weights). The obtained network KKn consists of 32,409 nodes and 2,799,530 edges. In the obtained network, the weight of the edges between the nodes (keywords) is equal to the *fractional* co-occurence of keywords i and j in the same works.

$$\mathbf{KKn} = n(\mathbf{WK})^T * n(\mathbf{WK})$$

where

$$n(\mathbf{WK})[w,k] = \frac{\mathbf{WK}[w,k]}{\max(1, \text{outdeg}(w))}$$

3.2 Networks of key words co-occurence

Exploratory analysis showed that in the obtained network the most frequentlty words *social*, *network*, and *analysis* were connecting most of the other keywords, that's why we excluded these 3 nodes from the network. Using Islands approach, we aimed to obtain subnetworks sized from 2 to 75 nodes. We got a large number of islands (342), where the majority of islands (301) represent just pairs of keywords. The main island includes 75 nodes; there are also some islands of smaller sizes.

Large part of the Main island (Figure 1) are the keywords on the topic of networking sites and social media (such as *networking, media, online, site, facebook, internet, technology, web 2.0*). Other central nodes are *information* associated with networking group, words *diffusion* and *privacy*, as well as *base* and *datum* (which also have links to many other keywords, including *big*, and *mining*). Other two central keywords are *model* and *graph*, which are connected to each other and other nodes, such as *dynamics, complex, spread, influence* (for the first one) and *random, theory, centrality – betweenness, large – scale*

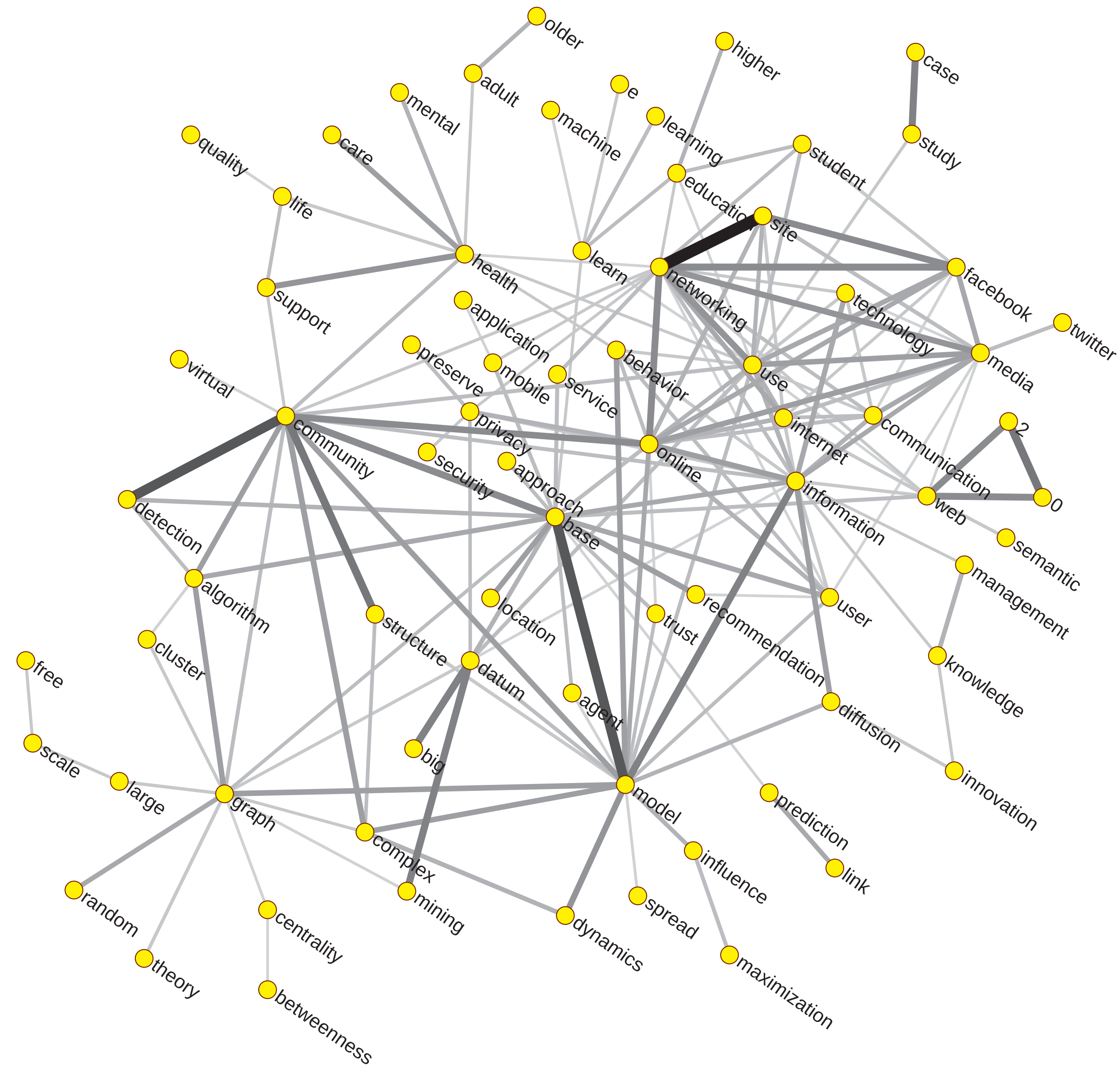

Figure 1: KK network Main Island

– *free, cluster* (for the second). These central nodes are also connected to the words *community* and *algorithm*, which have links to *detection* and *structure*. Other topics appeared in this subnetwork are associated with *health* and *education*.

Other islands (largest are represented at the Figure 2 identify some topics being studied in network analysis (*strength, weak, tie; corporate - interlock - directorate; triadic - closure; small - world*, or some broad topics under study (*organ - donor - donation; persecutory - delusion - paranoia; trade - international - migration*), as well as some stable phrases (*special, issue, introduction*).

Probably, we do not need the Pic? Not much sense

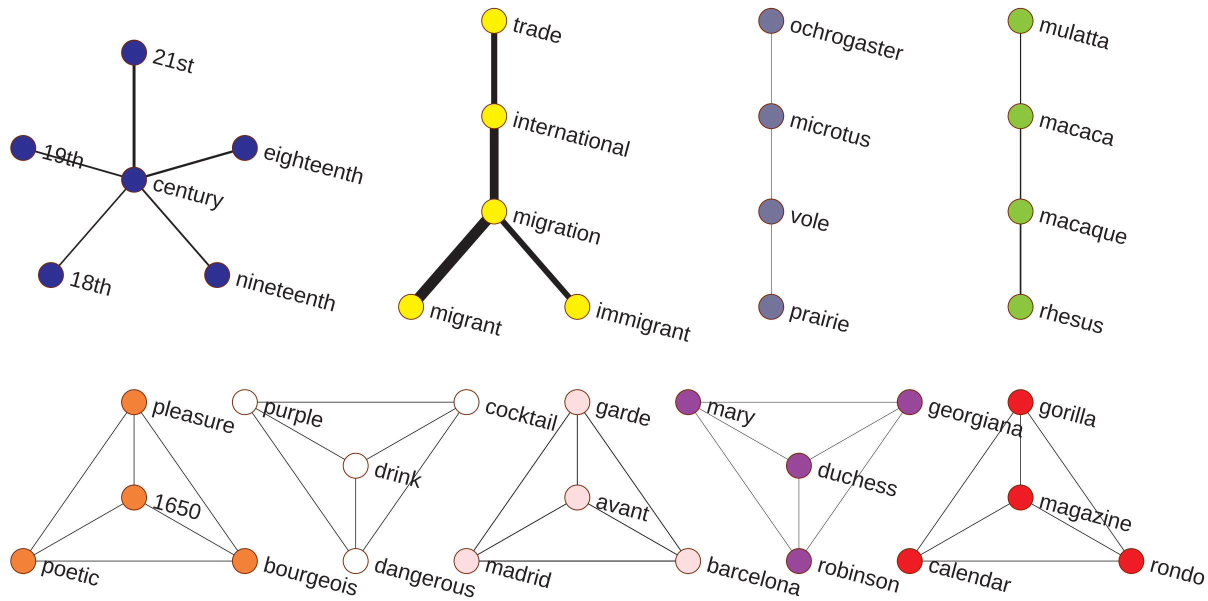

Figure 2: KK network Medium size Islands

4 Authors Collaboration

4.1 Networks creation

There are different ways to create one-mode networks of collaboration between authors AA out of two-mode networks of works and authors WA, which were presented and used in the previous studies (Batagelj and Cerinšek, 2013; Batagelj et al., 2014). Multiplying the reduced **WAr** network, consisted of 70,792 works and 93,011 authors, to transposed WAr network, and using different types of normalizations, we created three collaboration networks **Co**, **Cn**, and **Ct'**.

The standard and the easiest way to obtain the collaboration network is to create a **Collaboration network Co** (Batagelj and Cerinšek, 2013) by the multiplication of a transposed WAr network to original WAr network. In derived **Co** network, the weight of the edges between the nodes i and j is equal to total number of works author i and j wrote together. The loops in Co are equal to the total number of works that each author have (which is also equal to the indegree values of the WA network).

$$\mathbf{Co} = \mathbf{WA}^T * \mathbf{WA}$$

Collaboration network Cn uses **fractional approach**, where the contribution of authors to their own works and works written with co-authors is considered. The normalization creates network $n(WA)$ where the weight of each arc is divided by the sum of weights of all arcs having the same initial node as this arc (outdegree of a node). The network is constructed by the transposition of the WA network and multiplying it with the (normalized) $n(WA)$ network.

$$n(\mathbf{WA})[w, a] = \frac{\mathbf{WA}[w, a]}{\max(1, \mathrm{oudeg}[w])}$$

then

$$\mathbf{Cn} = \mathbf{WA}^T * n(\mathbf{WA})$$

In the derived network Cn, the weight of the edges between the nodes (authors) is equal to the contribution of author i to works, that he or she wrote together with author j (which can be not symmetric). Then the **author's total contribution** to all his/her works is counted as a value of diagonal (loops) of a Cn network for a certain author. Based on it, Batagelj and Cerinšek (Batagelj and Cerinšek, 2013) proposed **self-sufficiency index Si** as a proportion of author's contribution to all his/her works *cnii* and her/his total number of works (which is equal to the value of Indegree of author in WA network), and the **collaborativness index Ki**, which is complementary to it (is equal to 1 minus self-sufficiency).

$$S_i = \frac{cn_{ii}}{\text{indeg}_{\mathbf{WA}}(i)}$$

$$K_i = 1 - S_i$$

Using another type of normalization – Newman's normalization, who interprets collaboration in a "strict" way, as a collaboration only with others and not with author himself of herself, – the **Collaboration network Ct'** was constructed. In this case, for the initial WA network the weight of each arc is divided by the sum of weights of all arcs having the same initial node as this arc (outdegree of a node) subtracting the initial author (which is 1). Then the network **Ct'** is constructed by the transposition of the regularly normalized $n(WA)$ network (used for **Cn** network production above) and multiplying it with the Newman normalized $n'(WA)$ network.

$$n'(\mathbf{WA})[w,a] = \frac{\mathbf{WA}[w,a]}{\max(1, \text{outdeg}(w) - 1)}$$

then

$$\mathbf{Ct'} = n(\mathbf{WA})^T * n'(\mathbf{WA})$$

The obtained Ct' is undirected without loops. The contribution of a complete subgraph corresponding to each work is 1. The weights of the edges between the nodes (authors) are equal to the total contribution of "strict collaboration" of authors i and j to works they wrote together. The total contribution for an author is counted by line weights – it is equal to the sum of the weights of all the works he or she co-authored.

4.2 Collaboration between authors

It was already shown in the first paper that the authors having the largest number of papers have Chinese and Korean names. The issue of the super-productivity of these groups of authors was discussed by Harzing (2015) – this is the well-known "three Zhang, four Li" effect: 80% of people in China have one of only around 100 surnames. Thus, there is a high chance that different authors, having the same surname and first letter of the name, shrink together, creating 'multiple personalities'. This problem could be overcame if we would use a special ID (such as ORCID) for each scientist (but this information is not provided in WoS yet).

In this sense, it is not productive to look at the 'most writing' authors. However, from the **Co** network we still get an important information about collaboration between groups of authors in terms of amount of works written by two authors together. In the obtained **Co** network, we made a link cut at the level of at least 10 works written together, and got a subnetwork of 420 nodes, which includes 1 component of 58 nodes, 1 component of 9 nodes, 2 components of 7 nodes, 4 components of 6 nodes, 4 components of 5 nodes, 9 components of 4 nodes, and 23 components of 3 nodes. Almost half of the nodes (45%) belong to the 95 components of the size 2. There are not so many authors who has 10 works written in collaboration, and even for those who have it is quite common to have them written in pairs.

However, there are some extra cases. The component of 58 nodes is formed by the authors with Chinese and Korean names (Figure 3). Again, we can propose that this subgroup is a result of "multiple personalities": there still a higher chance that there are several (as an example) `ZHANG_Z` collaborationg with several `LI_Y`, whose names are merged together. The largest line weights are for `MA_J` and `WANG_Y`, who have 31 works written together. According to our database, the author with the name `Ma J.` has published works in different spheres such as computer science, social networks, energy, physics, or health. The second name of this author varies: it is *Jing, Jun, Jiemin*, etc. Name `WANG_Y` also varies: *Wang YZ, Wang YC, Wang YW*, etc. Other such pairs with large amount of papers in common are `WANG_B` and `WU_B` (27), `GUO_B` and `YU_Z` (25).

Several selected components of the size 3 and more are presented on Figure 4. Some of these structures are complete subgroups, where everyone is connected to everyone: group of Khadilkar, Kantarcioglu, Thuraisingham, Khan, Abrol, Heaterly, working in the field of online social networks and social media (minimum 22 works written together); McCarty, Killworth, Bernard, Johnsen, and Shelley, working in the field of methods for measuring personal social networks (minimum 11 works written togeter, with 22 works between Killworth and Bernard); Kimura, Saito, Ohara, and Motoda, working in the field of artificial intelligence (minimum 23 works); Lax, Buccafurri, Nicolazzo, Nocera connected to Ursino, working in the field of online social networks (minimum 11 works); Kennedy, Green, Golinelli, Wenzel, Tucker connected to Zhou, working in the filed of family, sex, social support (minimum 10 works). Several groups represent more star-like structures: a group around Latkin with minimum 10 works with Mandell, and maximum 27 works with Davey-Rothwell, in the field of psychology and medicine; and a group around Kazienko, having maximum 28 works with Bródka analyzing complex networks. Other groups represent authors working in the field of medical studies (Vassilev, Rogers, Kennedy, end others), ERGM (group around Robins and Pattison, having 38 works written together), animal and human bahaviour (James, Croft, Krause, and Wilson; Farine, Sheldon, and Firth), rsik and desease social networks (Potterat; Muth, Rothenberg), blockmodeling (Batagelj, Ferligoj, and Doreian). Among 95 pairs, there are those with maximum values (number of works written in brackets): Fowler and Cristakis (43), Carminat and Ferrari (32), Borgatti and Everett (29), and well-known pairs such as Valente and Fujimoto (16), Maybody and Rezvania (15), Dunbar and Roberts (13), Barabasi and Posfai (11), Brandes and Lerner (11), Litwin and Stoeckel (10).

It is interesting to compare the number of works that author has with the values of her or his contribution to these works and level of collaborativeness with others (Table 1). Because of the "multiple personalities" problem we had to exclude the names of the Chinese and Korean authors from the output. The names are ordered by authors' fractional total contribution to the field.The authors with the indexes of collaborativeness larger then 50% are marked in boldface.

In terms of **largest productivity**, the top rated authors are social scientist R. Burt, followed by phisician M. Newman, who have total contribution larger then 50. They are followed by P. Doreian, H. Park, and R. Dunbar, whose total contribution is larger then 40. Other authors with quite large total contribution values (from 30 up to 40) are B. Wellamn, T. Valente, S. Park, P. Bonachich, L. Leidersdorf, C. Latkin, H. Litwin, and P. Marsden. Among the authors in the table, several have a very **large total number of works** – C. Latkin (130), T.Valente (97), R.Dunbar (91), M.Newman (81). However, the **indexes of collaborativeness** varies across these authors – while it is quite high for R.Dunbar, T.Valente, and C.Latkin, for other mentioned authors with the large total contribution the level of collaborativeness is significantly lower. For example, for R.Burt, who has the highest total contribution and high total

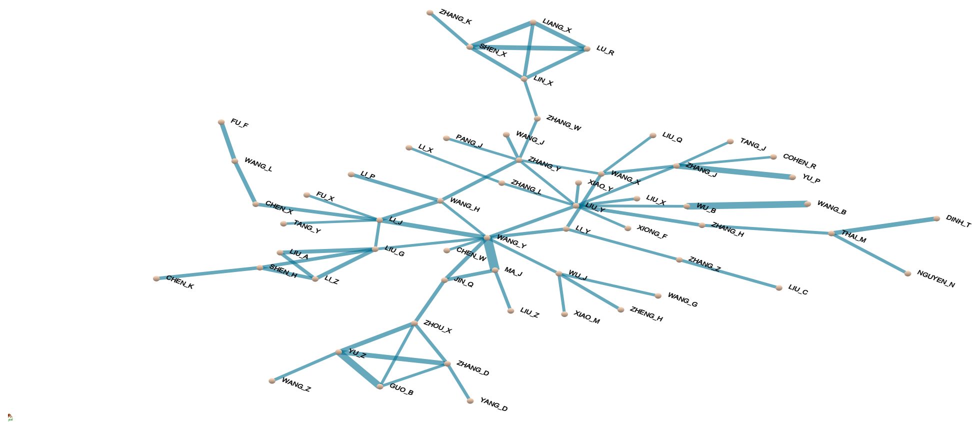

Figure 3: Co net: Main component

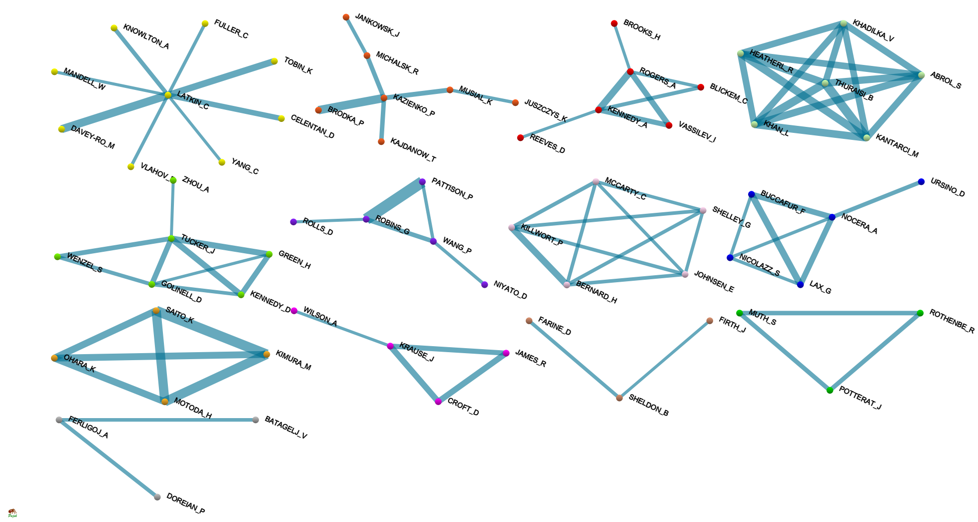

Figure 4: Co net: Selected components

Table 1: Collaborativeness

#	Author	Total contri-bution	Total # works	Collabora tiveness	#	Author	Total contri-bution	Total # works	Collabora tiveness
1	BURT_R	55,73	71	0,22	31	KRACKHAR_D	18,24	38	**0,52**
2	NEWMAN_M	50,02	*81	0,38	32	FALOUTSO_C	17,86	60	**0,70**
3	DOREIAN_P	46,19	72	0,36	33	JACKSON_M	17,78	38	**0,53**
4	DUNBAR_R	40,02	*91	**0,56**	34	GONZALEZ_M	17,76	52	**0,66**
5	WELLMAN_B	36,43	63	0,42	35	MOODY_J	17,7	40	**0,56**
6	VALENTE_T	34,96	*97	**0,64**	36	SCOTT_J	17,54	28	0,37
7	BONACICH_P	34	46	0,26	37	MORRIS_M	17,22	43	**0,60**
8	LEYDESDO_L	33,28	51	0,35	38	RODRIGUE_J	15,9	52	**0,69**
9	LATKIN_C	32,99	*130	**0,75**	39	WASSERMA_S	15,64	35	**0,55**
10	LITWIN_H	32,42	50	0,35	40	KLEINBER_J	15,05	34	**0,56**
11	MARSDEN_P	30,17	39	0,23	41	BATAGELJ_V	14,64	33	**0,56**
12	BORGATTI_S	29,72	71	**0,58**	42	WILLIAMS_A	14,5	31	**0,53**
13	SNIJDERS_T	29,63	67	**0,56**	43	SINGH_A	14,5	36	**0,60**
14	FRIEDKIN_N	28,17	36	0,22	44	BRANDES_U	14,39	35	**0,59**
15	CARLEY_K	28,11	72	**0,61**	45	BERKMAN_L	14,3	39	**0,63**
16	BARABASI_A	27,61	67	**0,59**	46	MASUDA_N	14,26	28	0,49
17	WHITE_H	27,28	42	0,35	47	SMITH_A	14,2	40	**0,65**
18	CHRISTAK_N	22,89	74	**0,69**	48	LAZEGA_E	14,17	26	0,46
19	EVERETT_M	22,58	44	0,49	49	CONTRACT_N	14,15	43	**0,67**
20	KAZIENKO_P	21,97	64	**0,66**	50	GONZALEZ_A	14,13	35	**0,60**
21	MARTINEZ_M	21,9	53	**0,59**	51	PENTLAND_A	14,12	41	**0,66**
22	JOHNSON_J	21,19	54	**0,61**	52	FARINE_D	14,04	34	**0,59**
23	FOWLER_J	20,14	65	**0,69**	53	SCHNEIDE_J	13,89	52	**0,73**
24	SKVORETZ_J	20,07	42	**0,52**	54	WATTS_D	13,67	27	0,49
25	FREEMAN_L	20,03	27	0,26	55	FAUST_K	13,5	25	0,46
26	BREIGER_R	19,73	31	0,36	56	SMITH_M	13,29	39	**0,66**
27	ROBINS_G	19,67	64	**0,69**	57	RODRIGUE_M	13,21	46	**0,71**
28	RAHMAN_M	19,18	59	**0,67**	58	RICE_E	13,09	48	**0,73**
29	PATTISON_P	18,94	58	**0,67**	59	BONACICH_P	34	46	0,26
30	THELWALL_M	18,41	37	**0,50**	60	CROFT_D	11,6	46	**0,75**

number of works, this index is only 22%. However, there are quite a lot of other scientists having the level of collaborativeness larger then 50%.

We used the **Ct'** network (constructed using Newman's notion of 'strict collaboration') to get the groups of authors stronger connected to each other. We used Islands approach [] (simple and general islands) for extraction of these subgroups. Setting different lower and upper bounds of the subgroups, different amount of subgroups can be generated.

Using **simple islands** approach, we generated 14,222 islands of the size between 2 and 50 nodes (which are 45,524 nodes, or 45% of all nodes in the network). Four largest islands consists of, respectively, 35, 23, 21, and 19 nodes; other 69 islands have between 12 and 18 nodes. The largest part of the network (78%) consists of clusters of relatively small sizes – 2 (28%), 3 (24%), 4 (15%), and 5 (10%). The variation of a upper and lower trasholds change the situation in the following way: with the treshold [5,50] we get 2,192 islands composed of 14,215 (15% of all nodes in the network); with the treshold [10,50] we get 173 islands composed of 2,064 nodes (2.2% of all nodes); and with the treshold [20,100] we get just 3 islands composed of 79 nodes (0.1% of all nodes), which sizes are 35, 23 and 21 nodes. We decided to use the treshold [2,50] for further analysis.

The sizes of islands shows that there are many groups of collaborating authors that can be extracted

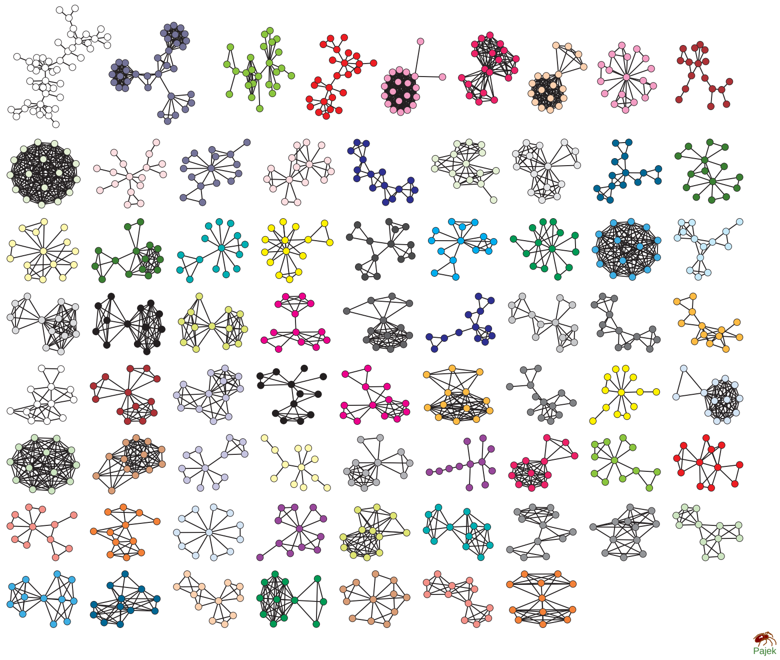

Figure 5: Ct' net: Islands 1-74

out of the **Ct'** network. There are different ways of how to identify the islands that can be interesting for the further inspection: to look at the 1) islands largest by size, 2) islands with largest values of line weights, 3) islands including some interestion, known names.

Using first approach, we extracted 74 **largest islands** with the size from 12 to 35 nodes (1,037 nodes, 2.2% of network), which are presented in Figure 5. Part of these structures are not very interesting: they are star-like networks, which represent one author collaborating with many others, or (almost) complete clusters (cliques), where all authors collaborate with (mostly) everyone else. However, islands having non-trivial structures can be intesting to inspect. The Figure 6 provides the pictures of 34 selected islands (501 nodes). Among them, the groups of physicists – Newman, Clauset, Girvan, Watts, Strogatz, Kossinets, Park, – and social gerontologists with Liwin in the middle – are identified; however, other authors included are not so well-known (as well-known authors can appear in smaller groups).

Another way to get interesting cases to inspect is to find the islands which have the **largest line weights between the nodes**. In largest islands, the ranges between smallest and largest line values are pretty low – they vary in the borders [0.018 – 2.00] for the island of 35 nodes, [0.035, 2.202] for the island of 23 nodes, [0.005, 1.00] for the island of 21 nodes, and [0.022, 0.330] for the island of 19 nodes. To get islands with really strong ties, we removed all the lines lower then certain trashold [7.5] from the **Ct' net** and got the network of 32 nodes. Then we manualy searched for the islands to which these 32 nodes belong, and extracted them. These islands are presented on Picture 7.

Some of the authors with the largest line weights between each other already appeared as the result of **Co** network line cut. These are the groups around Kimura, Saito, Ohara, and Motoda (artificial intelligence); Latkin and Davey-Rothwell (psychology and medicine); those who appeared as a pair: Borgatti and Everett (SNA methodology and UCINET), Fowler and Cristakis (health studies), Carminati and Ferrari (computer science), Barabasi and Posfai (physics), Litwin and Stoeckel (social support and

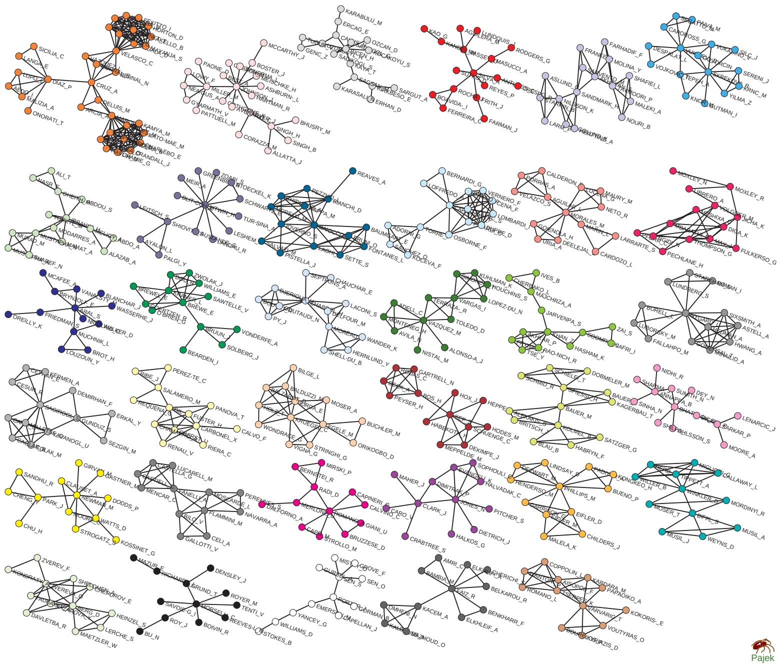

Figure 6: Ct' net: selected islands

health studies). There is also a group of authors working in medicine around Steinhausen and Metzke, physicists Grabowski and Kosiński working on artificial neural networks, representatives of urban studies Arentze and Timmermans. There are also several groups of authors with Chinese and Korean names.

Exclude Snijders from picture

Using third approach to get interesting islands for inspection, we seacrhed for the clusters to which **certain authors** belong.To make the story moe interesting, we decided to include to the list those authors who have a high level of collaborativeness according to **Cn** network (Table 1), as well some well-known figures in social network analysis. We manualy searched for the simple islands to which these 15 authors belong, and presented them in Picture 8. Groups around physicists Newman and Watts, as well as around Dunbar, Brandes, and Kazienko are quite large in comparison two others groups.

It is interesting to see how the groups obtained by the procedure of simple islands, where the authors are tightly connected to each other, can possibly form larger subnetworks. To find such structures, we used non-simple **general islands approach** and generated islands with different number of nodes: 13,200 islands of the size between 2 and 50 nodes (which are 47,991 nodes, or 52% of all nodes in the network), 2,630 islands of size [5,50] (20,396, or 22% of all nodes), 514 islands of the size [10,50] (7,374, or 8% of all nodes), and 70 islands of the size [20,100] (1,971, or 2.2% of nodes). It can be seen with this approach the amount of nodes included into the clusters in each treshold is larger then with the usage of simple islands.

Increasing the upper boundary of an island to 100 (and setting the lower boundary to 2, to be able to cath the pairs of strong collaborators), we finally got 13,182 islands with 48,029 nodes (22% of all nodes in network), with the largest islands of sizes of 96 and 80 nodes. The islands of the sizes of 20 nodes and more form only 4% of all nodes; largest part of the network (67%) consists of islands of size 2 (24%), 3 (21%), 4(13%) and 5 (9%). Two largest islands are presented in the Picture 9. The first island is formd of the authors with Chinese and Korean names; second cluster presents many authors conected to each

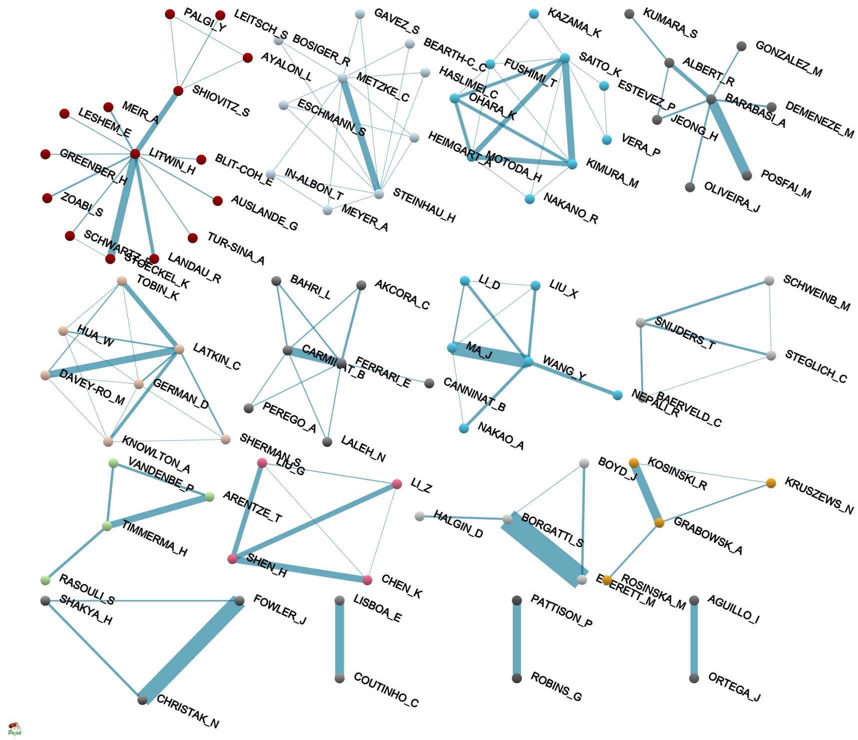

Figure 7: Ct' net: Simple islands for authors with largest line weights

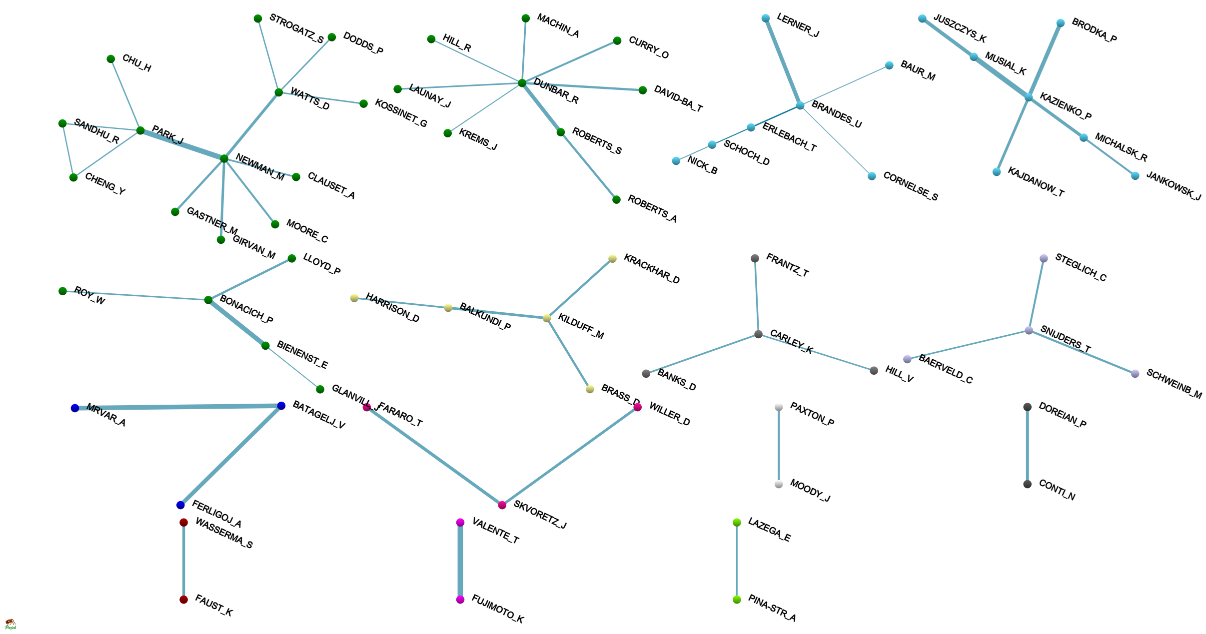

Figure 8: Ct' net: Simple islands for selected authors

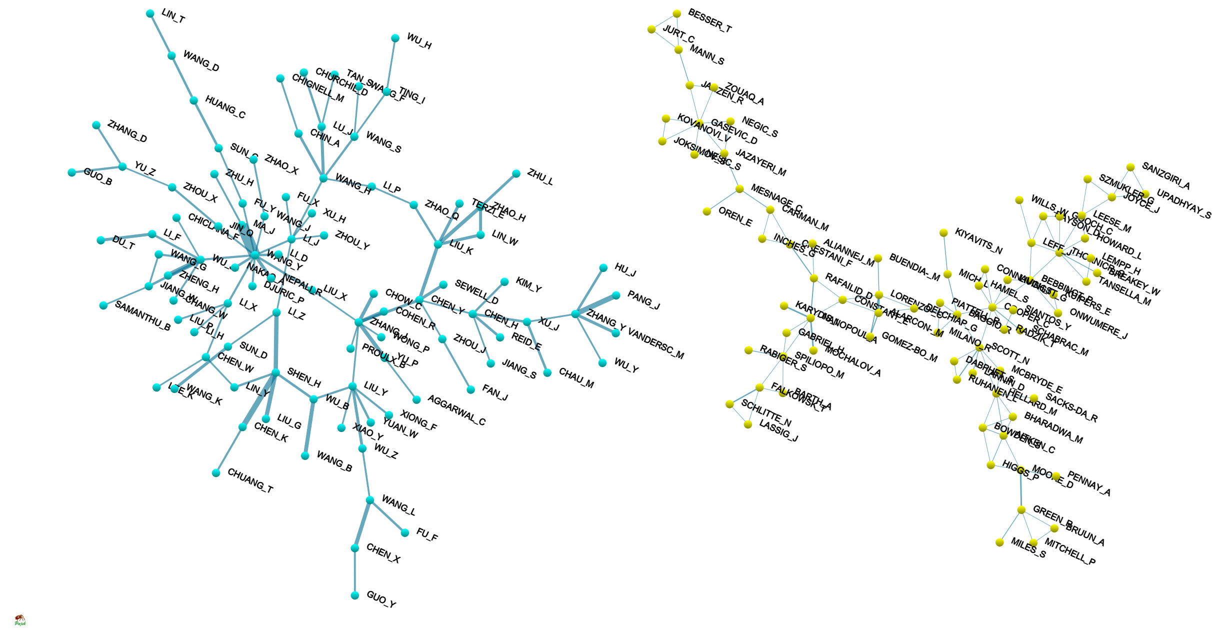

Figure 9: Ct' net: Largest regular islands

other – however, these are not the "core" participants in SNA field.

Again, as the inspection of all islands is time-consumimg and all results can not be presented in the article, we manually searched for some general islands, formed out of several simple islands, for certain well-known authors. The obtained islands are presented in the Figure 10. It can be seen that previous simple islands represented by Snijders, Skvoretz, Batagelj, Doreian, and Wasserman now form one connected island. Other two islands that were merged into one island are represented by Krackhard and Carley. The size of some clusters (represented by Dunbar, Valente, Moody) increased, while some islands (represented by Newman, Brandes, Kazienko, Bonachich).

To better know what topics the authors from the obtained islands are working at, we made an analysis of key words for some of the groups, which is presented in the following section.

5 Key words in coauthorship islands

5.1 Network creation

To construct the network of authors and keywords **AK**, we used normalized reduced networkds **WAr** and **WKr**. The first network was transposed and then multiplied with the second in the following way:

$$\mathbf{AK} = n(\mathbf{WA})^T * n(\mathbf{WK})$$

In the obtained network, the weight of the edges between the nodes a and k is equal to the fractional contribution for a given keyword k to the works of author a or a group of authors A.

$$\mathbf{AK}[\text{A,k}] = \sum_{a \in C} \mathbf{AK}[a, k]$$

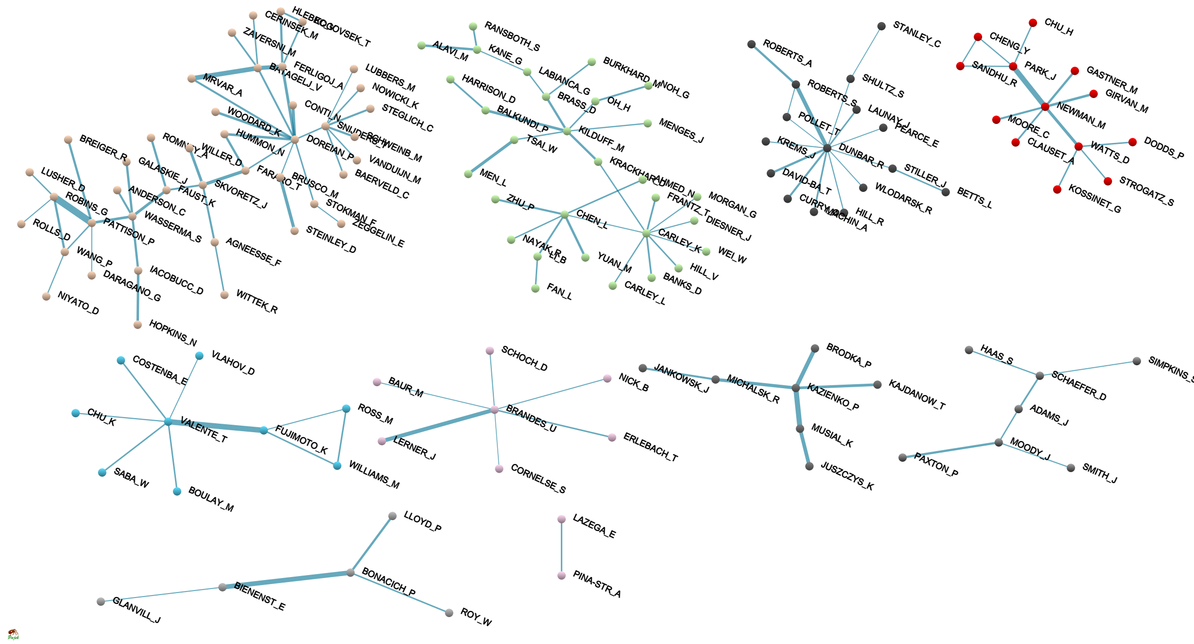

Figure 10: Ct' net: Regular islands for selected authors

5.2 Key words in selected islands

Based on the resluts obtained in the previous section, we have chosen the islands represented by BORGATTI_S, BARABASI_A, CHRISTAKIS_K, PATTISON_P, SNIJDERS_T, VALENTE_T for extraction of keywords. Top-20 keywords for each group are presented in the Table 2.

The top words for all these clusters, except the one represented by Skvoretz, are *network* and *social*, other keywords are special for each island. The island of Borgatti, Everett, Boyd, and Halgin is associated with the keywords *graph, centrality, role, regular, equivalence, semigroup, structure, clique, homomorphism*, and others. For the group of 8 authors with Barabasi, Posfai, Albert, etc. the keywords are *dynamics, complex, scale, web, community, world, internet, model, free, evolve, random*. The group of Fowler, Christakis, and Shakya have keywords *spread, behavior, health, smoking, human, cooperation, obesity, influence, evolution, dynamics*. Other group of social network analysts Robins and Pattison have the words *model, graph, random, markov, ligit, ligistic, regression,exponential, p, semigroup, asterisk, multirelational*. The group of Valente and is represented by the words *health, diffusion, behavior, innovation, peer, adolescent, influence, smoking, prevention, cigarette, leader*. On the official home page of Thomas Valente it can be found that he is working on the topics of *social networks, behavior change, and program evaluation* and *uses social network analysis, health communication, and mathematical models to implement and evaluate health promotion programs designed to prevent tobacco and substance abuse, unintended fertility, and STD/HIV infections*, and *also engaged in mapping community coalitions and collaborations to improve health care delivery and reduce healthcare disparities* [https://profiles.sc-ctsi.org/thomas.valente].

Table 2: Clusters of authors and keywords: Selected simple islands

	BORGATTI_S	BARABASI_A	CHRISTAKIS_K

Rank	Value	Id	Value	Id	Value	Id
1	4.9303	network	7.0709	network	3.1788	network
2	2.5918	social	2.0782	social	2.9358	social
3	2.0858	graph	1.7068	dynamics	1.0204	spread
4	1.4210	centrality	1.6670	complex	1.0192	behavior
5	1.4202	analysis	1.6362	scale	0.7261	health
6	1.3399	role	1.5946	web	0.5512	large
7	1.2780	regular	1.5516	community	0.5169	model
8	1.2424	equivalence	1.4709	world	0.4778	smoking
9	1.0530	semigroup	1.3622	internet	0.4522	human
10	1.0000	correction	1.1906	model	0.4479	cooperation
11	0.9891	structure	1.1858	free	0.4313	obesity
12	0.7755	clique	1.0210	evolve	0.4125	influence
13	0.7576	homomorphism	1.0087	science	0.3973	life
14	0.7241	relation	0.9808	random	0.3728	dynamics
15	0.6346	power	0.9476	wide	0.3715	evolution
16	0.6301	betweenness	0.8178	human	0.3463	analysis
17	0.6287	exchange	0.8076	theory	0.3286	cosponsorship
18	0.6232	algorithm	0.7561	small	0.3044	norm
19	0.6167	similarity	0.7536	graph	0.3036	trial
20	0.5595	ebloc	0.6603	phenomenon	0.2985	study
Total:	63.0810		76.6373		46.8865	
	PATTISON_P		SNIJDERS_T		VALENTE_T	

Rank	Value	Id	Value	Id	Value	Id
1	2.2196	network	2.6375	network	2.5536	network
2	2.0729	social	2.0902	social	1.9553	social
3	1.7567	model	1.6702	model	1.0000	untitled
4	1.3084	graph	1.0692	graph	0.9419	health
5	0.8939	random	0.8857	dynamics	0.8737	diffusion
6	0.8583	markov	0.7390	markov	0.7802	behavior
7	0.8531	logit	0.6903	random	0.7402	innovation
8	0.8220	logistic	0.6734	friendship	0.6974	model
9	0.8220	regression	0.6228	datum	0.6521	use
10	0.8012	exponential	0.5932	statistical	0.6349	peer
11	0.7055	analysis	0.5780	behavior	0.6216	adolescent
12	0.6752	p	0.5547	analysis	0.5717	influence
13	0.5530	statistical	0.5423	peer	0.5610	smoking
14	0.5038	structure	0.5383	inference	0.5371	analysis
15	0.3561	semigroup	0.5346	influence	0.5247	prevention
16	0.3522	asterisk	0.4623	stochastic	0.4987	cigarette
17	0.3368	process	0.4612	actor	0.4979	opinion
18	0.3333	multirelational	0.4480	selection	0.4860	leader
19	0.3249	family	0.4372	longitudinal	0.4545	risk
20	0.3031	dynamics	0.3785	orient	0.4491	intervention
Total:	38.6110		46.6732		44.8812	
	SKVORETZ_J		WASSERMAN_S		DOREIAN_P	

Rank	Value	Id	Value	Id	Value	Id
1	3.8058	network	2.4529	network	6.0097	network
2	1.6586	power	1.6875	social	3.7088	social
3	1.6277	exchange	1.0000	correction	1.5308	equivalence
4	1.5218	social	0.9414	analysis	1.4972	evolution
5	1.2301	bias	0.7270	model	1.4917	journal
6	1.0751	model	0.5509	graph	1.2177	structural
7	1.0000	correction	0.4818	datum	1.0395	measure
8	0.9204	structure	0.4595	method	0.9402	structure
9	0.7765	theory	0.4457	exchange	0.8107	group
10	0.6341	theorem	0.4319	stochastic	0.7987	balance
11	0.5001	tie	0.4282	structure	0.6923	analysis
12	0.4119	structural	0.3554	statistical	0.5395	actor
13	0.3972	weak	0.3501	blockmodel	0.5067	blockmodeling
14	0.3905	approximation	0.3438	kinship	0.4917	utility
15	0.3905	simulation	0.3308	equivalence	0.4870	model
16	0.3883	dynamic	0.3118	structural	0.4711	generalized
17	0.3436	theoretical	0.3079	logit	0.4667	stand
18	0.3371	strength	0.2666	relation	0.4339	connectivity
19	0.3108	analysis	0.2611	triad	0.4333	ranking
20	0.3105	sociology	0.2611	census	0.4238	regular
Total	33.5190		29.1417		48.1875	

The group of 4 authors connected through Snijders has the keywords *markov, random, friendship, behavior, peer, inference, influence, stochastic, actor, longitudinal, orient* which of represent their work in stochastic actor-oriented models. The island represented by Skvoretz has the keywords *power, exchange, bias, model, correction, theorem, approximation, simulation, dynamic*, and others. The pair of Wasserman and Faust can be represented with the words *correction, model, exchange, stochastic, structure, statistical, blockmodel, equivalence, logit, triad*, etc. (there are also *logistic* and *regression* on 23 and 24 places), and pair of Doreian and Conti – the words *equivalence, evolution, journal, balance, blockmodeling, generalized, regular, ranking*. The keywords for the regular island that these simple islands formed are presented in Table 3. We can see that the words with largest values are more commonly used words, such as *network, social, model, analysis, graph, structure, datum, structural, theory, group, method*, however, there are also special words which were already mentioned in the islands above, such as *correction, exchange, equivalence, random, power, markov, evolution, statistical, dynamics, generalized, regression, exponential, blockmodell, logit, p, cluster, ligistic, dynamic, blocmodeling*, etc.

Thus, we can conclude that the proposed method allows to find the keywords for sertain authors structures and represents the works of the mentioned authors.

6 Citation among authors

6.1 Network creation

To get information about citations among authors we computed the **CiteA** network as a product of multiplication of the networks **WAr** and **CiteR**.In this network, the value of the element $CiteA[u, v]$ is equal

Table 3: Clusters of authors and keywords: Selected general island

Rank	Value	Id	Rank	Value	Id
1	30.0225	network	21	1.8844	generalized
2	20.1127	social	22	1.8226	journal
3	8.6241	model	23	1.8012	regression
4	7.3574	analysis	24	1.7816	exponential
5	6.0054	graph	25	1.7772	blockmodel
6	5.5047	structure	26	1.7639	logit
7	3.1894	datum	27	1.7326	balance
8	3.0265	structural	28	1.7253	p
9	3.0000	correction	29	1.6844	measure
10	2.9594	exchange	30	1.6639	algorithm
11	2.7971	equivalence	31	1.6584	cluster
12	2.6809	random	32	1.6381	approach
13	2.5432	theory	33	1.6222	actor
14	2.5255	power	34	1.5873	logistic
15	2.5081	markov	35	1.5509	relation
16	2.4107	evolution	36	1.5398	introduction
17	2.2839	group	37	1.5356	bias
18	2.2531	statistical	38	1.5144	dynamic
19	2.1939	method	39	1.4467	blockmodeling
20	2.1816	dynamics	40	1.4391	friendship
Total				371.7399	

to the **number of citations** from works coauthored by u to works coauthored by v.

$$\mathbf{CiteA} = (\mathbf{WA})^T * \mathbf{Cite} * \mathbf{WA}$$

Using fractional approach, we also produced the normalized version of this network **CiteAn**. Normalization creates network **n(CiteR)** where the weight of each arc is divided by the sum of weights of all arcs having the same initial node as this arc (outdegree of a node). The value of element $CiteAn[u, v]$ is equal to the **fractional contribution** of citations given by author u to author v.

$$\mathbf{CiteAn} = (\mathbf{WA})^T * n(\mathbf{Cite}) * \mathbf{WA}$$

where

$$n(\mathbf{Cite})[u, v] = \frac{\mathbf{Cite}[u, v]}{\max(1, \text{outdeg}(u))}$$

6.2 Citations

Using CiteA network, we looked at the loops and got a list of athors with largest **absolute values** of self-citation (Table 4). The authors with the highest absolute value of self-citations are R.Dunbar, having 589 citations, and Latkin with 387 citation (but also having the largest total number of works, as it was seen in the previous section). Other authors with large number of self-citing (more then 200 times) – Christakis, Valente, Burt, Newman, Robins, Pattison, Fowler, and Barabasi – also have quite a high total number of works. Interestingly, Croft and James, from animal social networks, do have respectively high level of self-citations, but the tital number of works fro them, respectively, 46 and 38.

After extraction of loops from CiteA network, we looked at the authors with the **largest lines weights** (largest times of authors citing other authors). We used line cut (with treshold 95) and got the networks of 43 nodes, with several not connected components (Figure 11). One of these groups is star-like group of physicists with Newman in the middle, with most of the nodes citing him, and him citing Barabasi and Albert (withthe first also citing the second), and Watts. Another groups is represented by social scientists, with Robins and Pattison connected to each other, as well as to Wasserman: Robins is also citing Handcock and Snijders (who is being cited by Steglich); three of them are also cited by Lomi. There is also a group of Fowler and Christakis; Latkin in the center citing others, Roberts and Dunbar citing each other (and Dunbar citing Hill), group of authors working in the field of animal social networks (Croft, Krause, James citing each other and Farine citing James and Sheldon), and some others.

To overcome the overrepresentation of the authors with many works, we used fractional approach. Looking at the **loops** of the network **CiteAn**, we can see the highest **fractional values** of self citations (Table 5, column "Value", again, authors with Chinese and Korean names were excluded). We also counted **the proportion** of authors' *self-citations to themselves* and their *external citations to others* by dividing the vector of loops to the vector of outdegree in the CiteAn network (Table 5, column "%"). The proportion of self-citation according to the number of all other citations varies for different authors, however, it is larger for Burt (24%), Killworth (19%), Dunbar (18%), Wellman (17%), Leydesdorf (17%), and Atzori (17%).

We used Islands approach (general) to get islands of the size [5, 200] from the normalized network **CiteAn** and got 414 islands of size between 5 and 200 (3,418 nodes, or 3.7% of all the nodes in network). Almost half of these islands (48%) are of sizes 5 (18%), 6 (16%), and 7 (14%). The largest island Main island of **CiteAn** network consits of 200 nodes (6%). We tried to move the upper border of the treshold up to 500 nodes, and it resulted with the Main island of 500 nodes, which means that there is large

Table 4: Authors with the highest self-citation: absolute values

Rank	Id	Value	Rank	Id	Value
1	DUNBAR_R	589	11	BARABASI_A	201
2	LATKIN_C	387	12	FARINE_D	191
3	CHRISTAK_N	292	13	SNIJDERS_T	188
4	VALENTE_T	280	14	WELLMAN_B	153
5	BURT_R	268	15	DOREIAN_P	148
6	NEWMAN_M	248	16	BORGATTI_S	146
7	ROBINS_G	232	17	ZENOU_Y	146
8	PATTISON_P	224	18	RICE_E	143
9	FOWLER_J	221	19	JAMES_R	142
10	CROFT_D	204	20	KRAUSE_J	141

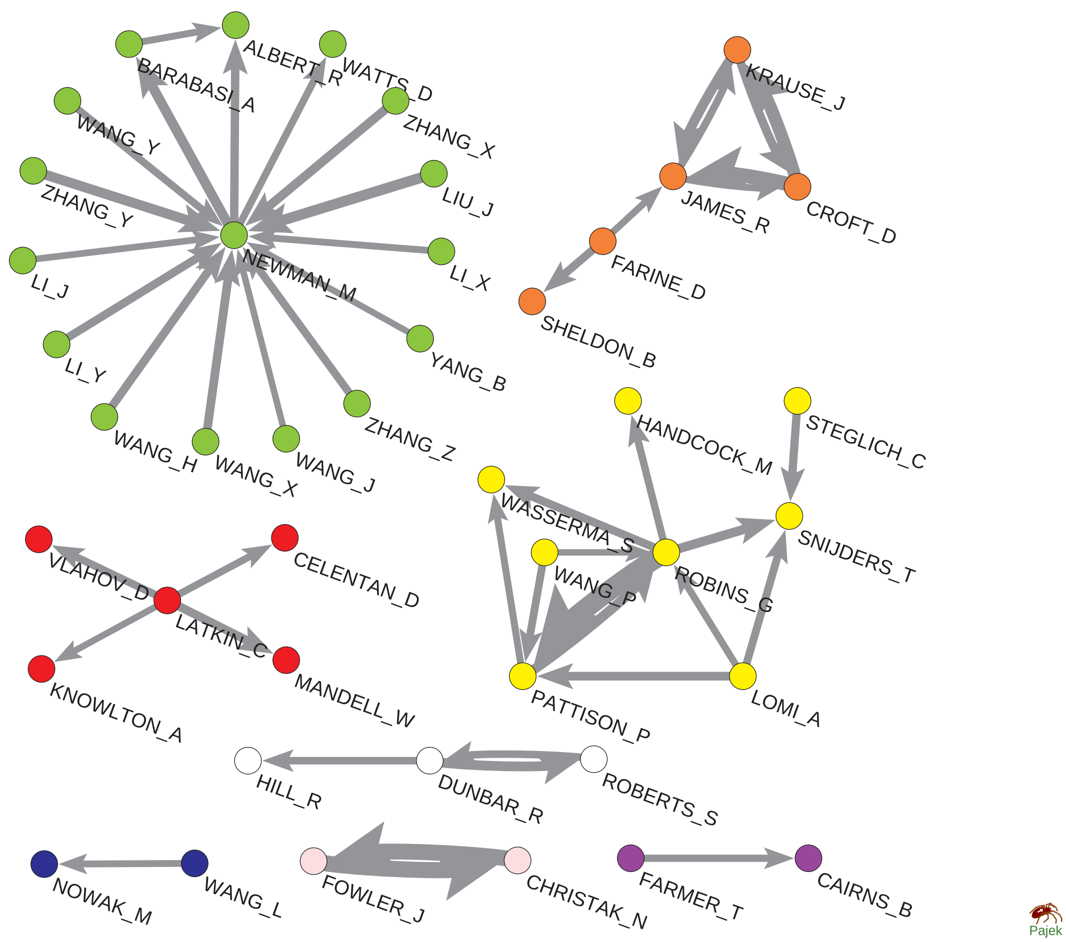

Figure 11: CiteA network: Largest line weights

Table 5: Authors with the highest self-citation

#	Author	loops	outdegree	%	#	Author	loops	outdegree	%
1	DUNBAR_R	49.34	274.31	**0.18**	21	FARMER_T	11.88	96.06	0.12
2	LATKIN_C	43.23	508.88	0.08	22	CROFT_D	11.82	151.10	0.08
3	JUNG_J	29.26	210.06	0.14	23	KILLWORT_P	11.00	59.19	**0.19**
4	BURT_R	25.52	104.54	**0.24**	24	BERNARD_H	10.97	68.68	0.16
5	CHRISTAK_N	24.45	180.29	0.14	25	JAMES_R	10.50	109.82	0.10
6	VALENTE_T	24.27	248.25	0.10	26	FRIEDMAN_S	10.35	211.26	0.05
7	BARABASI_A	21.48	143.84	0.15	27	ATZORI_L	10.15	59.75	**0.17**
8	NEWMAN_M	20.23	173.70	0.12	28	FARINE_D	9.78	115.99	0.08
9	WELLMAN_B	17.08	99.91	**0.17**	29	EVERETT_M	9.34	69.52	0.13
10	DOREIAN_P	16.48	121.24	0.14	30	KRAUSE_J	9.03	99.98	0.09
11	FOWLER_J	15.99	156.65	0.10	31	BERKMAN_L	8.95	88.85	0.10
12	SNIJDERS_T	13.88	140.77	0.10	32	POTTERAT_J	8.90	88.58	0.10
13	LEYDESDO_L	13.80	78.93	**0.17**	33	MORRIS_M	8.86	110.99	0.08
14	ROTHENBE_R	13.46	145.26	0.09	34	KAZIENKO_P	8.80	151.34	0.06
15	RICE_E	12.77	163.03	0.08	35	TUCKER_J	8.50	210.81	0.04
16	CARLEY_K	12.59	152.26	0.08	36	SKVORETZ_J	8.48	79.38	0.11
17	PATTISON_P	12.57	138.92	0.09	37	BATAGELJ_V	8.42	62.48	0.13
18	KELLY_J	12.19	162.15	0.08	38	MUTH_S	8.38	125.57	0.07
19	BORGATTI_S	11.98	125.75	0.10	39	MARTINEZ_M	8.31	103.10	0.08
20	ROBINS_G	11.96	162.65	0.07	40	LITWIN_H	8.30	97.33	0.09

nested group of nodes. Main island of 200 nodes is presented on the Figure 12. In the center there is a group of physicists, surrounded by tje cloud of citations from the 'Chinese group'. Most of these authors are again representatived of already mentioned 'multiple authors' – Harzing (2015) found them to be highly presented in the field Physics. For the purposes of better visualization, the authors cited Newman, Barabasi separately and both of them together are put to the same place (which produce 3 black lines in the picture). This group also inclues Strogatz, Leskovec, Girvan, Evans. Left upper part of the graph is consisted of the authors from the Social sciences, traditional representatives of Social network analysis. Among them, the authors with the strongest ties are Borgatti and Everett; the authors having the largest numbders of incoming ties are Freeman (7) is Burt (4); Batagelj also has 2 connections from Ferligoj and Doreian. There is also a trird component of authors positioned around Wasserman, who is being cited by the authors also connected to each other – Robins and Pattison, Snijders, Holland. Another author citing Wasserman (and Faust) largerly is Kazienko, also having incoming citations (they were also seen in the structures before). Another local star having a lot of incoming ties is Dunbar, having reciprocal connections with Roberts and Pearce.

Second and third largest islands are composed of 66 and 41 nodes, respectively. Other islands have 26 and less nodes. Some selected large islands are presented on the Figure 13. First and second islands (if we count from left to right by three lines) represented by Migliano and Marasco, and Kaskutas, respectively, contain researchers from the fields of medicine, health, desease and addiction. The island of Latkin, Friedman, Potterat, and others was already mentioned above – this is a group in the field of psychology and health. The group of Chrstakis and Moreno (different ones) represents the researchers working in the field of adolescents, teens in social media, internet addiction, and health. Island formed with Farmer contains researchers working with behavioral and emotional disorders, psychology and youth. Next group represented by Atzori is the one working in the field of internet of things, the one of Folke – ecology. There is a group of scientists around French sociologist Bourdieu. The group of animal

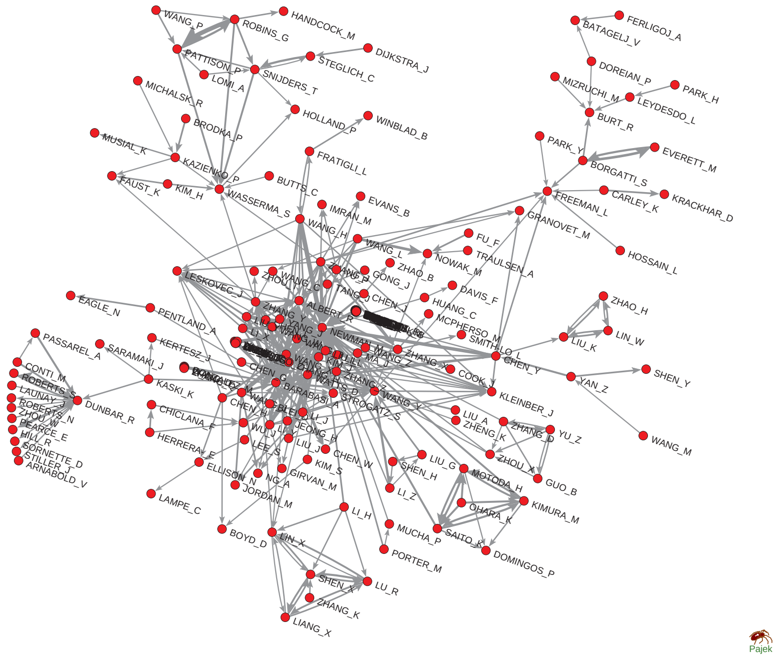

Figure 12: CiteA:Main island

social network analysts Croft, Krause, Farine, and others is quite well-connected. We do not have space and time to drill into all the exported islands, but we can see that there is a large number of groups with different topic specialization.

7 Citation among journals

7.1 Networks creation

To get information about citations among journals, we computed the **CiteJ** network, which takes into account citations from papers published in journal *i* to papers published in journal *j*, which appeared in the works included into the **WJr** network (the analysis was limited to networks with complete descriptions of works). We used a **CiteR** network to get information on citations between works. In the obtained network, the value of the element $CiteJ[u, v]$ is equal to the **number of citations** from journal i to journal j.

$$\mathbf{CiteJ} = (\mathbf{WJ})^T * \mathbf{Cite} * \mathbf{WJ}$$

As journals of different sizes were included into the data set, using the *fractional approach* the normalized $n(CiteR)$ network was also used to produce a **CiteJn**. The value of the element $CiteJn[u, v]$ is equal to the **fractional contribution** of citations from papers published in journal i to papers published in journal j.

$$\mathbf{CiteJn} = (\mathbf{WJ})^T * n(\mathbf{Cite}) * \mathbf{WJ}$$

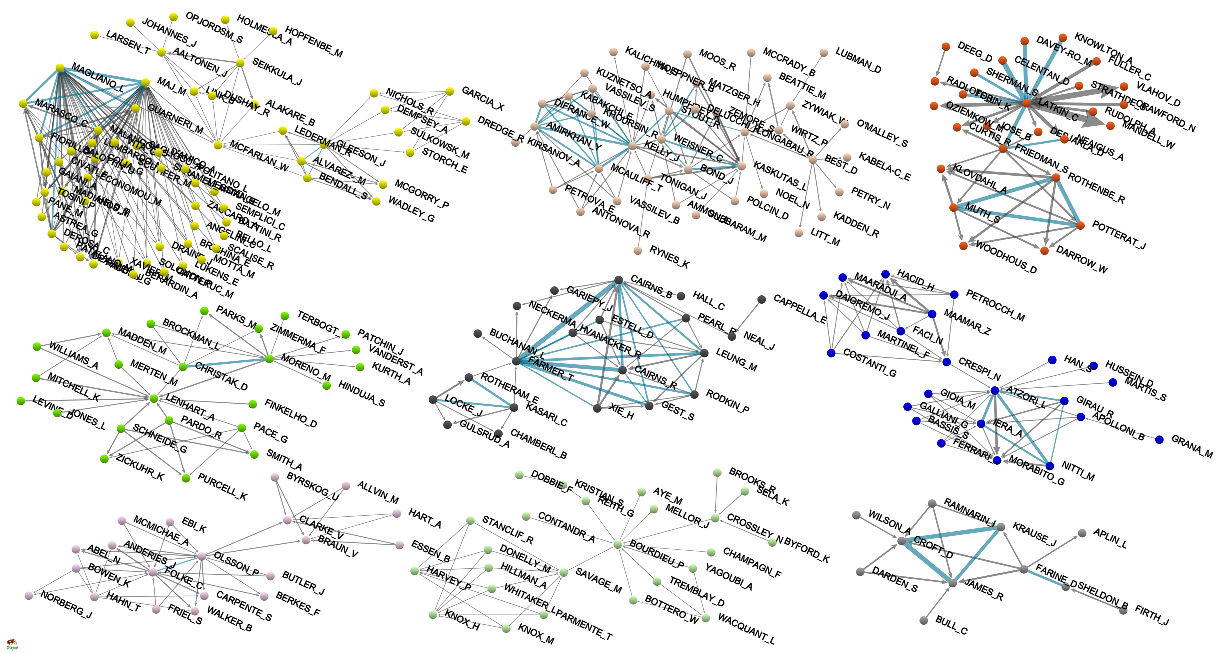

Figure 13: CiteA network: Largest line weights

7.2 Citations

The loops of the **JJs** network show the journals with the highest **absolute values** of self citations (Table 6). They are *Social Networks* with more then 4,000 citations, and *Computers in Human Behavior* – with twice less. *Physica A – Statistical Mechanics and its Applications*, *Physical Review E*, *Lecture notes in Computer Science*, *Cyberpsychology, Behavior, and Social Networking*, and more social sciences oriented *Social Science & Medicine* and *American Journal of Sociology'* have also relatively high absolute values of citations.

After removing loops from the same network, we used a line cut (with a treshold of 205) and extracted the subnetwork of 27 journals with the largest amounts of absolute values of citations given to each other (Figure 14). The largest component of this network consists of two parts, with *Social Networks* in the center, being largerly cited by the journals from Social sciences, and also citing back the *American Journal of Sociology* and *Journal of Mathematical Sociology*. However, it is also cited a lot from the journals of the second part of the component, belonging to the fields of Natural sciences and general scientific journals – *Plos One, Lecture Notes in Computer Science, Physica A*. The journals in this part are quite connected with each other. The journal *Computers in Human Behavior* form a separate group, connected to *Cyberpsychology, Behavior, and Social Networking* (reciprocally), as well as to *Journal of Computer-Mediated Communications* and *Personality and Individual Differences*. Another pair of journals are the *Animal Behaviour* citing *Behavioral Ecology and Sociobiology*.

The loops of the **JJsfn** network show the journals with the highest **fractional values** of self-citations (Table 7, column "Value"). Even though some level of self-citation is is typical for all journals, there are some journals that have larger levels. Again, the highest value belongs to the *Social Networks* journal. Other highly ranked journals are from the fields of Computer science and Cyberpsychology – the journals *Computers in Human Behavior* and *Lecture Notes in Computer Science* (the previous order is a bit changed). Other journals with relatively high number of internal citations are already mentioned before

Table 6: Journals with the highest self-citation: absolute values

Rank	Value	Id
1	4443	SOC NETWORKS
2	2058	COMPUT HUM BEHAV
3	569	PHYSICA A
4	429	PHYS REV E
5	382	LECT NOTES COMPUT SC
6	339	CYBERPSYCHOL BEHAV
7	328	SOC SCI MED
8	315	AM J SOCIOL
9	303	PLOS ONE
10	258	ANIM BEHAV
11	246	SCIENTOMETRICS
12	232	J MED INTERNET RES
13	226	P NATL ACAD SCI USA
14	209	ORGAN SCI
15	194	BEHAV ECOL SOCIOBIOL

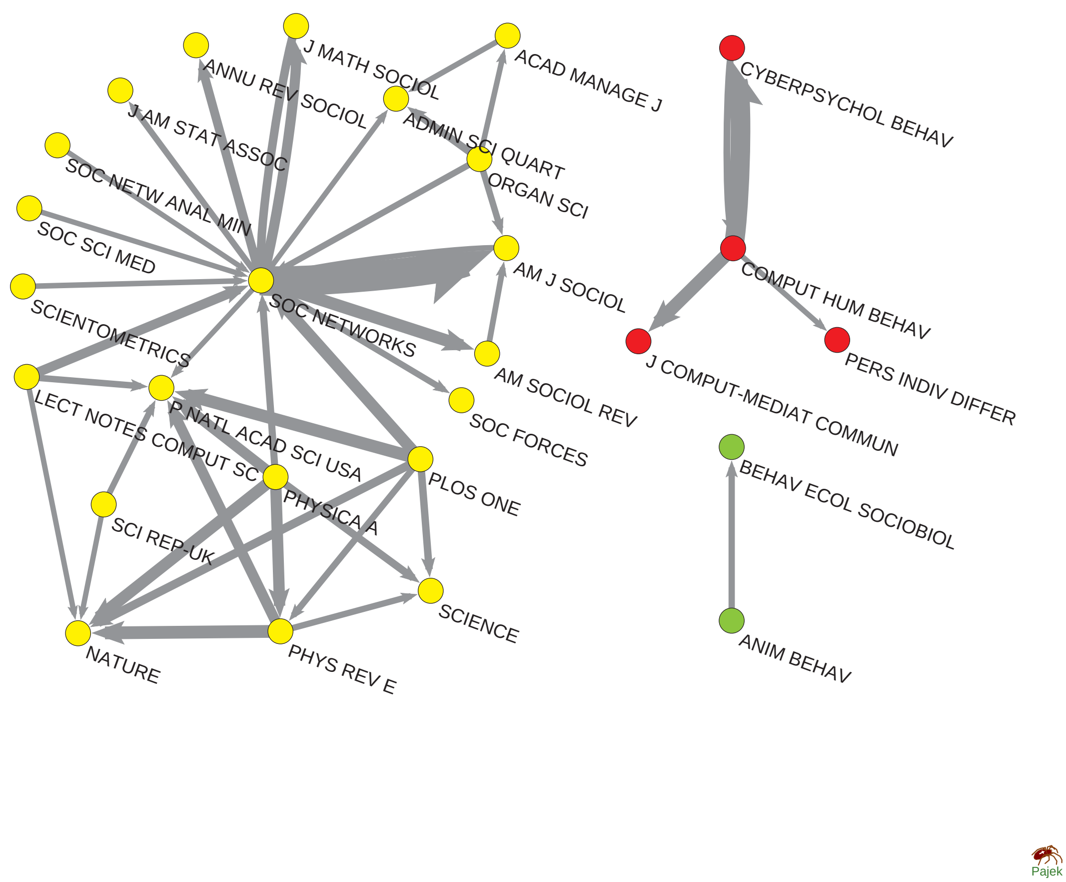

Figure 14: JCiJ network: Largest line weights

Table 7: Journals with the highest self-citation

#	Value	%	Journal	#	Value	%	Journal
1	**355.65**	**0.34**	SOC NETWORKS	16	18.35	0.17	ANIM BEHAV
2	**168.39**	0.22	COMPUT HUM BEHAV	17	17.03	0.12	AIDS BEHAV
3	**122.57**	0.09	LECT NOTES COMPUT SC	18	16.03	0.19	AM J COMMUN PSYCHO
4	57.75	0.13	PHYSICA A	19	14.87	0.10	INFORM SCI
5	43.00	0.14	SOC SCI MED	20	14.14	0.14	KNOWL-BASED SYST
6	42.18	**0.24**	J MED INTERNET RES	21	12.64	0.19	PROF INFORM
7	41.49	0.21	CYBERPSYCHOL BEHAV	22	12.35	**0.23**	COMUNICAR
8	33.16	0.05	PLOS ONE	23	12.00	0.18	BEHAV ECOL SOCIOBI
9	32.93	0.11	PHYS REV E	24	11.87	**0.25**	AM J EPIDEMIOL
10	30.22	0.13	SCIENTOMETRICS	25	11.01	0.11	DECIS SUPPORT SYST
11	24.16	0.14	P NATL ACAD SCI USA	26	10.58	0.14	J ETHN MIGR STUD
12	23.15	**0.26**	AM J SOCIOL	27	10.43	0.13	COMPUT EDUC
13	20.04	0.05	LECT NOTES ARTIF INT	28	10.31	0.18	SEX TRANSM DIS
14	19.31	0.12	EXPERT SYST APPL	29	10.19	**0.28**	NATURE
15	18.77	0.14	NEW MEDIA SOC	30	9.85	0.09	ORGAN SCI

Physica A, Social Science & Medicine, *Cyberpsychology, Behavior, and Social Networking*, as well as Journal of Medical Internet Research, which appears in the table above on a lower place. The differences between values for the first three listed journals and others are quite significant.

However, as different journals have different number of works, we also counted **the proportion** of journals' *internal citations to themselves* and their *external citations to other journals* by dividing the vector of loops to the vector of outdegree in the JJsfn network (Table 7, column "%"). Several journals with the highest values – *Nature, American Journal of Sociology, American Journal of Epidemiology, Journal of Medical Internet Research, Comunicar Journal* – have quite high level of internal citation – around the quarter. *Social Networks* journal has even more – 34% of internal citations. Large level of citation within a journal to the same journal may mean that it is seen as an important source of information for the scientists involved into the particular field.

We generated Islands of size between 2 and 50, and got 195 islands (448 nodes of all the network, or 5.5%), with the largest island containing 50 nodes (10%), and 67% being just pairs of nodes. The Main islands is presented on the Figure 15.Citations among journals have clear acyclic (hierarchical) organization. There can be sevral main groups of journals detected: journals in Social Sciences (on the right), Computer Science (on the left), Physics (in the middle) and General scientific journals (in the bottom).

In the **Computer Science group** of journals, the most citing position is taken by *Lecture Notes in Computer Science*, which largerly cite the journals from all the groups: Computer Science (*Computer Networks, Communications of the ACM, Computers in Human Behavior*, and others), Social sciences (*Social Networks, Social Network Analysis and Mining, Structural Analysis in the Social Sciences, Annual Review of Sociology*), Physics (*Physical Review*, and General scientific journals (*Nature, Journal of Interdisciplinary Sciences, Proceedings of the National Academy of Sciences USA*. It also have equal number of outgoing and incoming citations with the journal *Lecture Notes in Artficial Intelligence*, which also cites a lot of journals from different fields. Another journal from this group – *Computers in Human Behavior* – is largerly citing the journal *Cyberpsychology, Behavior, and Social Networking*, which also cites it back, *Journal of Computer-Mediated Communication*, and other journals related to Behavioral studies and Information Systems, as well as Media.

The group of journals from the **Physics** is presented by only three journals – *Physica A – Statistical Mechanics and Its Applications, Physical Review E, Physical Review Letter*, with the first one citing others. The *Physica A, Physical Review E* are also the journals that cite largerly general scientific journals *Nature* and *Science*. It is also seen that the traditional, older **general scientific journals** – *Nature, Science, Journal of Interdisciplinary Sciences, Proceedings of the National Academy of Sciences USA* are cited by the newly emerged one – *Plos One*.

In the **Social Sciences group**, the most citing journal is *Social Networks*, which is strongly connected to the *Amercian Journal of Sociology*, as well as have outgoing connections to many other sociological journals (*Journal of Mathematical Sociology, Social Forces, American Sociological Review, Social Network Analysis and Mining*, which also cite it back (however, in a less degree), and *Annual Review of Sociology, Sociological Methodology*. This *Social Networks* journal is also cited by other journals from differnt fields of Social sciences (Sociology, Organizational science, Information science, Methodology), as well as from the journals of the two mentioned above groups. For this journal, the largest incoming citatons go from *Lecture Nodes of Computer Science, Lecture Notes in Artficial Intelligence, Plos One, Physica A*. Thus, the journal itself cites mostly journals from the field of Social sciences, but it is being cited by the journals from other disciplines – Computer, Physics, and General scientific fields. There are also links from *Lecture Notes in Computer Science* and *Plos One* to the *American Journal of Sociology*.

Other 30 obtained islands of journals citing each other sized from 3 to 6 (110 nodes) are the ones on the topics of: psychology and deviation; psychiatry; medicine; surgery; health disabilities; substance abuse and addiction; nursering; social work; archeology and anthropology; language and sociolinguistics; economics and economic behavior; education; conflicts and peacekeeping; library science; ergonomics; transportation. There are also three journals from the field of behavioral ecology and animal behavior: *Behavioral Ecology and Sociobiology* being cited by *Animal Behaviour* and *Behavioral Ecology*.

Other islands 32-195 contain (328 nodes) only 2 nodes – they are pairs of journals. Journals with the largest weights of lines are presented on the Table 8. The links are directed: first written journal cites second one. These journals cover the wide range of topics, including *health, health policy, psychology and psychiatry, adolescence, sex, STD and AIDS, migration, communication, demography, business and management, consuming behavior and marketing, information science, computing, language, peace and conflicts, engineering*.

8 Bibliographic Coupling

8.1 Networks creation

Bibliographic coupling occurs when two works each cite a third work in their bibliographies – and this suggests some content communality between these two works. Having more works citing pairs of prior works increases the likelihood of them sharing content [Batagelj, chapter 2]. We used **CiteR** network to produce a **Jaccard biCo** network, which can be determined as:

$$\mathbf{biCo} = \mathbf{Ci} * (\mathbf{Ci})^T$$

$$\mathbf{biCo_{pq}} = \text{\# of works cited by both works p and q} = \mid Ci(p)...Ci(q) \mid$$

Bibliographic coupling weights are symmetric: biCopq = biCoqp:

Table 8: Pairs of journals

#	value	from journal	to journal
1	8,1	IEEE GLOB COMM CONF	IEEE INFOCOM SER
2	6,26	HIST COMUN SOC	COMUNICAR
3	4,63	J YOUTH ADOLESCENCE	J RES ADOLESCENCE
4	4,44	INT J GERIATR PSYCH	J PSYCHIAT RES
5	4,38	INT MIGR	INT MIGR REV
6	4,31	J BUS ETHICS	ACAD MANAGE REV
7	3,99	DEMOGR RES	DEMOGRAPHY
8	3	J INTELL FUZZY SYST	J APPL MATHE COMPUT
9	3	J INT DEV	TROP MED INT HEALTH
10	3	PERVASIVE MOB COMPUT	INT CONF PERVAS COMP
11	2,78	J CONSTR ENG M	J CONSTR ENG M ASCE
12	2,68	PHYS EDUC RES CONF	PHYS REV SPEC TOP-PH
13	2,59	ENERGY RES SOC SCI	ENERG POLICY
14	2,5	INT P ECON DEV RES	TECHNOVATION
15	2,37	COMPUT ASSIST LANG L	LANG LEARN TECHNOL
16	2,33	INFORM SOC-ESTUD	PERSPECT CIENC INF
17	2,33	WORLD DEV	ECON J
18	2,31	J PEACE RES	J CONFLICT RESOLUT
19	2,22	HEALTH RES POLICY SY	HEALTH POLICY PLANN
20	2,1	SEX HEALTH	INT J STD AIDS
21	2	REV LAT COMUN SOC	PALABRA CLAVE
22	2	J RETAIL CONSUM SERV	AUSTRALAS MARK J
23	2	ETHN DIS	HEART LUNG
24	2	IEEE INT SYMP INFO	IEEE T INFORM THEORY
25	2	REV BRAS ENFERM	REV LAT-AM ENFERM

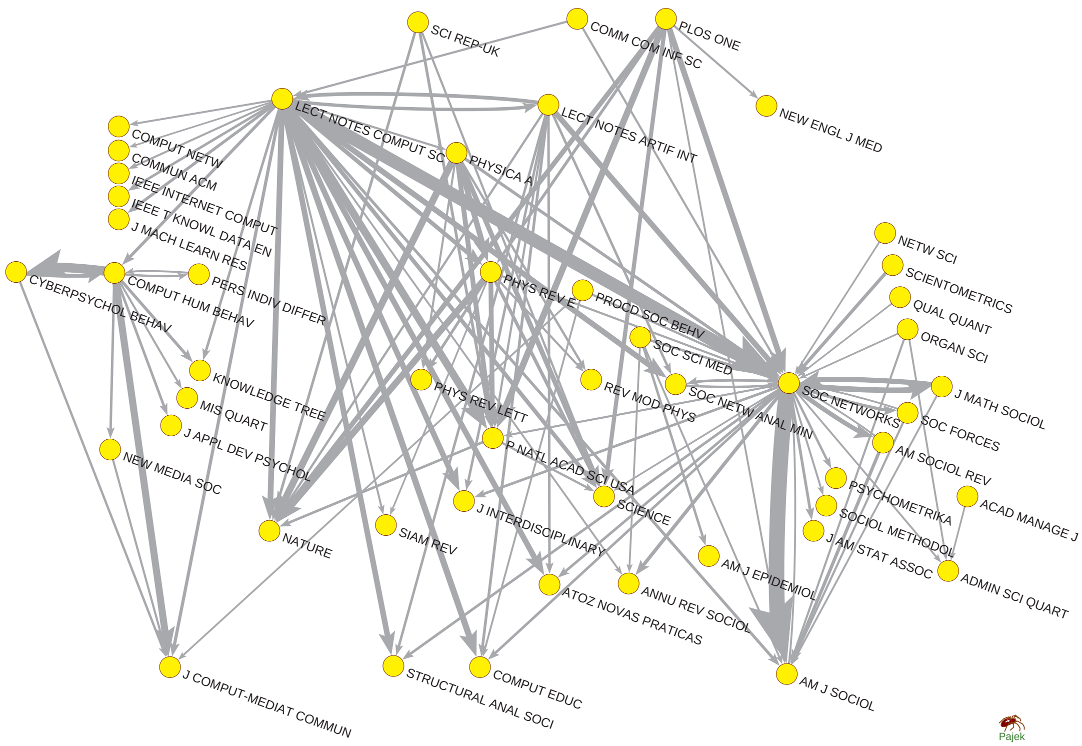

Figure 15: Citations among journals – Main island

$$\mathbf{biCo}^T = (\mathbf{Ci} * \mathbf{Ci}^T)^T = \mathbf{Ci} * \mathbf{Ci}^T = \mathbf{biCo}$$

Using obtained network, we constructed two networks – co-citations between journals **JCoj** and co-citations between authors **ACoj** – by its multiplication with normalized **WJsr** and **WAsr** networks (we limited our analysis to networks with complete descriptions of works). Normalization creates networks $n(WJr)$ and $n(WAr)$ where the weight of each arc is divided by the sum of weights of all arcs having the same initial node (journal or author) as this arc (outdegree of a node). Weights in the obtained networks take into account *fractional* similarity of journals i and j, or authors u and v.

$$\mathbf{JCoj} = n(\mathbf{WJ})^T * \mathbf{biCoj} * n(\mathbf{WJ})$$

$$\mathbf{ACoj} = n(\mathbf{WA})^T * \mathbf{biCoj} * n(\mathbf{WA})$$

The values of links from biCoj are redistributed in JCoj and ACoj. The total sum of link weights is preserved.

$$\sum_{e \in E(\mathbf{JCoj})} \mathbf{JCoj}[e] = \sum_{e \in E(\mathbf{biCoj})} \mathbf{biCoj}[e]$$

We should note that the produced Jaccard network **biCo** contained a large number of links – 62.079.457, – and that‘s why the computation of networks **JCoj** and **ACoj** would be quite time consuming. That‘s

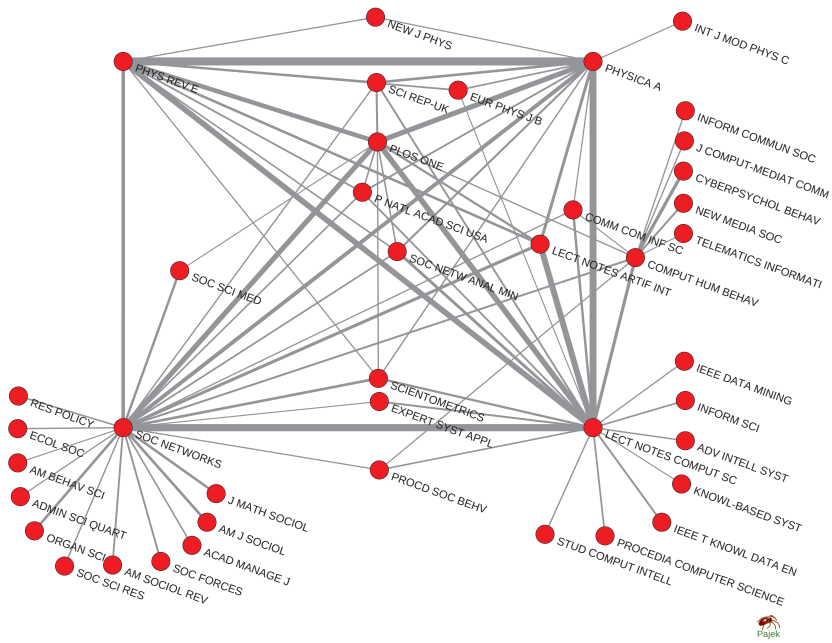

Figure 16: Cj network: Journals

change: full Jaccard network size, obtained networks sizes

why the desision was made to make a link cut in the Jaccard **biCo** at the level 0.085, and then use the obtained network for the multiplication with **WJsr** and **WAsr** networks, respectively. The obtained **biCo** network contained 70.792 works and 13.208.451 links (change). After multiplication, we got **JCoj** network with 8.943 nodes and 4.966.617 arcs (including XXX loops), and **ACoj** network with 93.011 nodes and 127.220.243 arcs (including 14.131 loops). In both obtained networks, the loops were deleted, the bidirected arcs were converted to edges (with summation of values) before the further analysis.

8.2 Co-citation among journals

After simplification, **JCoj** network contained 8,943 nodes and 5,136,616 edges. However, the majority of lines was of a very low value, and after the line cut at the level 1200 (as journals have great weights, a large link-cut was required), we got a network of 41 nodes.

Most of the journals presented on the Figure 16 are journals from the fields of Physics and Computer science closely connected to each other, which means – sharing the large amount of literature in their citations. Four such journals are particularly prominent: *Physica A, Physical Review E, Lecture Notes in Computer Science* and *Lecture notes in artificial intelligence*, with *Plos One*, representing general scientific journals. These results are similar to the ones obtained for the analysis of literature on clustering [Batagelj, Chapter 2]. *Social Networks* journal is also presented, having connections (and similar citation patterns) with all journals mentioned above as well as with journals from the group of *Social Sciences – Journal of Mathematical Sociology, Social Forces, American Sociological Review, American journal of Sociology, Social Science & Medicine* and *Scientometrics*. It does not have conections to *Social Network Analysis and Mining* journal – despite its name, it is focused more on data mining in large networks and reflects more a computer science orientation.

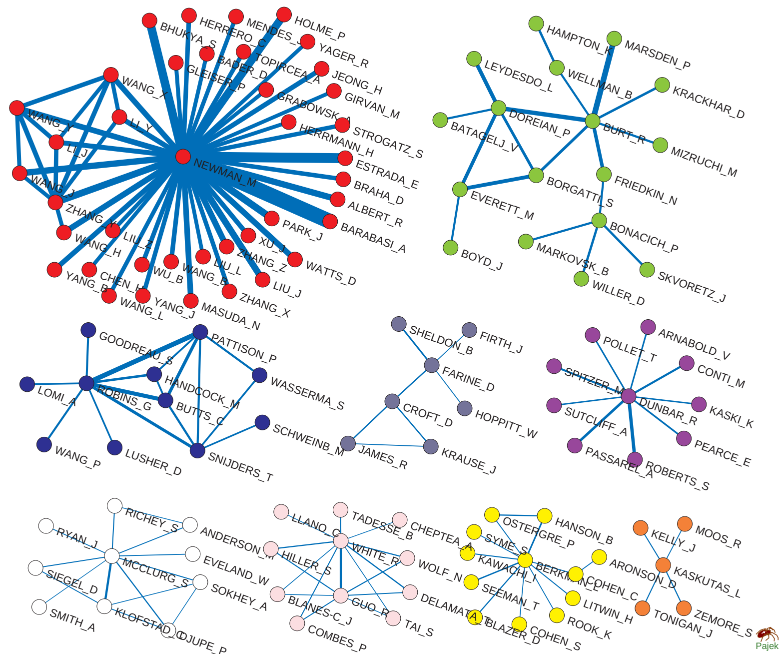

Figure 17: Jaccard network: Authors

8.3 Co-citation among authors

To make the obtained **ACoj** network manageble for the analysis, the line cut at the level 2.5 was done at first, and the network of 1.162 nodes and 9.926 arcs (including 402 loops) was obtained. After simplification, we got the network with same number of nodes and 4.762 of edges. We used Islands approach (regular islands) and extracted 9 islands of sizes from 5 to 40. These islands are presented in the Figure 17.

The larger island on the left comes from the physics driven literature and is centered on Newman. Most of the authors in this island are again Chinese authors, but it also includes such well-known physicists as *Barabasi, Albert, Watts* and *Strogatz*. The second and third islands contains sets of traditional social network scientists. In the second island, among 17 scientists the largest indegree weights are taken by *Burt, Doreian, Bonachich, Everett, Borgatti*. In the third island, the most central are *Robins, Pattison* and *Snijders*. The last island shows the similarity in citation patterns between authors from the field of animal social network analysis. There are five more islands of star-like structures, with *Berkman, Dunbar, Kaskutas, McClurg, R.White* in the center.

9 Conclusions

References

Barabâsi, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., Vicsek, T. (2002). Evolution of the social network of scientific collaborations. Physica A: Statistical mechanics and its applications, 311(3-4), 590-614.

Batagelj V. (2005). SN5 – network data for Viazards session at INSNA Sunbelt 2008. http://vlado.fmf.uni-lj.si/pub/networks/data/WoS/SN5.zip

Batagelj, V. (2017) WoS2Pajek. Networks from Web of Science. Version 1.5 (2017). URL: http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek:wos2pajek

Batagelj V., Cerinšek M.(2013). On bibliographic networks. Scientometrics. 96 (3), 845-864

Batagelj, V. (2014) Efficient Algorithms for Citation Network Analysis. arXiv:cs/0309023

Batagelj, V., Doreian P., V., Ferligoj, A., Kejžar N. (2014). Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution.

Batagelj, V., Ferligoj, A. Squazzoni, F. (2017) The emergence of a field: a network analysis of research on peer review. Scientometrics, 113: 503. https://doi.org/10.1007/s11192-017-2522-8

Batagelj V., Ferligoj A., Doreian P. (2019). Bibliometric analysis of the network clustering literature. In Doreian P., Batagelj V., Ferligoj A. Advances in network clustering and blockmodeling. Wiley.

Bonacich, P. (2004). The invasion of the physicists. Social Networks, 26, 285-288

Borgatti, S. P., Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. Journal of management, 29(6), 991-1013.

Brandes, U., Pich, C. (2011). Explorative visualization of citation patterns in social network research. Journal of social structure, 12(8), 1-19.

Carley, K. M., Hummon, N. P., Harty, M. (1993). Scientific influence: An analysis of the main path structure in the Journal of Conflict Resolution. Knowledge, 14(4), 417-447.

Clarivate Analytics (2018). URL: https://clarivate.com/products/web-of-science/databases/

Chinchilla-Rodríguez, Z., Ferligoj, A., Miguel, S., Kronegger, L., Moya-Anegon, F. (2012). Blockmodeling of co-authorship networks in library and information science in Argentina: A case study. Scientometrics. 93. 699-717. 10.1007/s11192-012-0794-6.

Cugmas M., Ferligoj A., Kronegger L. (2016). The stability of coauthorship structures. Scientometrics. 106(1), 163–186.10.1007/s1119201517904

Elsevier. Historical depth (2018). URL: https://www.elsevier.com/solutions/scopus/how-scopus-works/content

Ferligoj A., Kronegger L., Mali F., Snijders T., Doreian P. Scientific collaboration dynamics in a national scientific system. Scientometrics.

Franceschet, M. (2009). A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. Scientometrics, 83(1), 243-258.

Freeman, L. (2004). The development of social network analysis. A Study in the Sociology of Science, 1.

Freeman, L. C. (2011). The development of social network analysis–with an emphasis on recent events. The SAGE handbook of social network analysis, 21(3), 26-39.

Glaenzel, W., Andreas, S. (2004). Analysing scientific networks through co-authorship, in: Henk F., Moed et al., ed., Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems, Dordrecht, Boston, London, Kluwer Academic Publishers, 257-276.

Groenewegen, P., Hellsten, I., Leydesdorff, L. (2015) Social Networks as a looking glass on the social networks community. International Sunbelt XXXV Conference. Hilton Metropole, Brighton, UK, June 23 – 28, 2015. Abstracts, 118.

Harzing, A. W. (2015). Health warning: might contain multiple personalities—the problem of homonyms in Thomson Reuters Essential Science Indicators. Scientometrics, 105(3), 2259-2270.

Harzing, A. W., Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. Scientometrics, 106(2), 787-804.

Hilbert, F., Barth, J., Gremm, J., Gros, D., Haiter, J., Henkel, M., ... Stock, W. G. (2015). Coverage of academic citation databases compared with coverage of scientific social media: Personal publication lists as calibration parameters. Online Information Review, 39(2), 255-264.

Hou, H., Kretschmer, H., Liu, Z. (2008). The Structure of Scientific Collaboration Networks in Scientometrics. Scientometrics, 75,189. 10.1007/s11192-007-1771-3

Hummon N.P., Doreian P.(1989). Connectivity in a citation network: The development of DNA theory. Social Networks, 11, 1, 39-63.

Hummon N.P., Doreian P., Freeman L.C. (1990). Analyzing the Structure of the Centrality-Productivity Literature Created Between 1948 and 1979 / Science Communication. 11, 4, 459 – 480.

Hummon, N. P., Carley, K. (1993). Social networks as normal science. Social networks, 15(1), 71-106.

Hunter, L., Leahey E. (2008). Collaborative Research in Sociology: Trends and Contributing Factors. American Sociologists, 39, 4, 290-306.

Kejžar, N., Černe, S. K., Batagelj, V. (2010). Network analysis of works on clustering and classification from web of science. In Classification as a Tool for Research, 525-536. Springer, Berlin, Heidelberg.

Kronegger L., Mali F., Ferligoj A., Doreian P. (2012). Collaboration structures in Slovenian scientific communities. Scientometrics, 90, 631-647. 10.1007/s11192-011-0493-8.

Lazer, D., Mergel, I., and Friedman, A. (2009). Co-citation of prominent social network articles in sociology journals: The evolving canon. Connections, 29(1):43:64.

Leydesdorff L., Schank T., Scharnhorst A., De Nooy W. (2008). Animating the development of Social Networks over time using a dynamic extension of multidimensional scaling / El Profesional de Informacion, 17(6).

Lopaciuk-Gonczaryk, B. (2016). Collaboration strategies for publishing articles in international journals – A study of Polish scientists in economics. Social Networks, 44, 50-63.

Mali, F., KroneggerL., Ferligoj, A. (2010). Co-authorship trends and collaboration patterns in the Slovenian sociological community. Corvinus journal of sociology and social policy, 1(2), 29-50.

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: a systematic comparison of citations in 252 subject categories. Journal of Informetrics, 12(4), 1160-1177.

Moody, J. (2004). The structure of a social science collaboration network. American Sociological Review, 69, 213–238

Newman, M.E.J. (2004). Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences of the United States of America 101(Suppl1), 5200–5205.

Newman, M.E.J. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical Review E, 64.

Otte, E., Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. Journal of information Science, 28(6), 441-453.

Pontille, D. (2003). Authorship Parctices and Institutional Contexts in Sociology: Elements for a Comparison of the United States and France, Science, Technology and Human Values, 28, 2, 217-234.

Šubelj, L., Fiala, D. (2017). Publication boost in Web of Science journals and its effect on citation distributions. Journal of the Association for Information Science and Technology, 68(4), 1018-1023.

Sokolov M.M., Safronova, M.A., Guba, K.S., Dimke, D.V. Intellectual landscape and social structure of the local academic community (the case of St. Petersburg sociology) // Preprint.Humanities research. WP6. Higher School of Economics, 2012. No. WP6 / 2012/01.

Varga, A. V., Nemeslaki, A. (2012) Do organizational network studies constitute a cohesive communicative field? Mapping the citation context of organizational network research. Journal of Sociology and Social Anthropology, 5(64), XV, 349-364.

Wagner, C. S., Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. Research policy, 34(10), 1608-1618.

A Appendix